

# White Paper

## FUJITSU Server PRIMERGY & PRIMEQUEST

### Performance Report PCIe SSDs P3700

This document looks closely at the PCIe-SSDs from the P3700 series that are available for some PRIMERGY and PRIMEQUEST servers with regard to the disk I/O performance that can be achieved. In addition to the measurement results, the measurement method and the environment, in which the measurements were carried out, are also briefly explained.

**Version**

1.0b

2016-03-24



## Contents

Document history.....	2
Introduction and technical data.....	3
Measurement method.....	5
Measurement environment.....	6
Measurement results .....	7
A single PCIe-SSD.....	7
Two PCIe-SSDs.....	10
Four PCIe-SSDs .....	12
Storage capacity share under load .....	14
Access of several processes to a PCIe-SSD.....	16
Best practice .....	17
Optimizing the current processor frequency .....	17
Comparison with other storage media.....	19
Conclusion .....	20
Literature.....	21
Contact .....	21

## Document history

### **Version 1.0**

First report version

### **Version 1.0a**

Minor corrections

### **Version 1.0b**

Minor corrections

## Introduction and technical data

The PCIe-SSD is an innovative, nonvolatile storage medium for PRIMERGY and PRIMEQUEST servers, which in certain cases can be used as a logical hard disk drive instead of “Hard Disk Drives” (HDDs) or “Solid State Drives” (SSDs) that have a SAS or SATA interface. The fact that the PCIe-SSD storage medium is directly coupled with the PCIe bus means that very high transaction rates and low latencies are possible. Such a storage medium is interesting for access acceleration to data that is stored on a nonvolatile basis, in cases such as:

- Databases
- Web2.0 Servers

■ Page files (especially in virtual environments, because these files are an essential part of the memory management concepts there) Also new in the family of storage media discussed below is the form factor. The PCIe-SSDs still exist as a version for PCIe slots in servers. In addition, there are now also versions that can be conveniently connected externally in the 2.5” HDD bays of the servers explicitly provided for this purpose.

### Technical data

The PCIe-SSDs on offer are flash memories that are based on a multi-level cell (MLC) NAND type of memory. For a period of five years they allow 10 DDPD (drive writes per day) and thus belong to the write-endurance class “Mainstream Endurance”<sup>1</sup>. At present, there are three storage capacities, which – depending on the server model – can in each case be ordered in one of two form factors.

Storage capacity	Order codes	
	SFF versions (Small Form Factor = 2.5”)	AIC versions (Add in Card)
800 GB	SSD PCIe3 800GB Main 2.5' H-P EP	PCIe-SSD 800GB P3700
1.6 TB	SSD PCIe3 1.6TB Main 2.5' H-P EP	PCIe-SSD 1.6TB P3700
2 TB	SSD PCIe3 2TB Main 2.5' H-P EP	PCIe-SSD 2TB P3700

The two form factors do not differ in their performance.

A powerful controller is integrated on the storage medium itself, thus making an additional RAID controller unnecessary. The bus interface to the server is of type PCIe 3.0, x4, which would allow a practical throughput of up to about 3380 MB/s. The SFF versions are hot-pluggable. A special PCIe switch, which occupies a PCIe 3.0 x16 slot, is used for these versions in the RX/TX servers for every four PCIe-SSDs.

The thermal properties play a larger role for this generation of PCIe-SSDs than for most other storage media. In the case of load profiles with a very high write load the PCIe-SSDs considered here generate waste heat of up to 25 W. Due to the specific cooling situation in 2.5” HDD bays this heat emission is of special significance for the SFF versions. The storage medium itself ensures through timely throttling (with reduced performance) that its temperature does not exceed the limit of 71 °C. To avoid such thermal reductions in performance it is essential to observe the configuration and environment conditions, for which the servers and storage medium have been released. This means in particular:

- The most recently released version must be used for the iRMC firmware of the server. This is the only way to ensure that the fan regulation is prepared for the PCIe-SSDs.
- Changes to the released status of the server, which affect the air flow, are forbidden (e. g. removal of dummy covers).
- The ambient conditions may not impede the cooling (e. g. too high ambient temperature, obstructions to supply or extract air, etc.)

The PCIe-SSDs considered in this document are the third generation of this storage medium for PRIMERGY and PRIMEQUEST servers. The second generation (ioDrive<sup>®</sup>2 PCIe-SSDs) was covered in the white paper [“Performance Report PCIe-SSDs ioDrive<sup>®</sup>2”](#) under performance aspects.

<sup>1</sup> A detailed explanation of the special features of SSDs is available in the white paper [Solid State Drives - FAQ](#)

### Release matrix

The allocation of server models to orderable PCIe-SSDs, as depicted in the following table, is valid at the time of publication of this document. The PCIe-SSDs considered here are not released as drives for booting unless indicated by a note in the following table.

Servers	AIC	SFF
	#PCIe-SSDs P3700	#PCIe-SSDs P3700
PRIMERGY BX2560 M1	-	2 <sup>*)</sup>
PRIMERGY CX2550 M1	-	2
PRIMERGY CX2570 M1	-	2
PRIMEQUEST 2x00x2	8 - 32	-
PRIMERGY RX2530 M1	-	4
PRIMERGY RX2540 M1	-	8
PRIMERGY RX2560 M1	-	8
PRIMERGY RX4770 M1	-	4
PRIMERGY RX4770 M2	-	4
PRIMERGY TX2560 M1	-	8

*\*) Released for booting in UEFI mode with specific operating systems*

Some components may not be available in all countries or sales regions.

### Administration software

You can install the “Intel® Solid-State Drive Data Center Tool” and then access it with the command “isdct”, thus enabling you to monitor a great many features of the PCIe-SSDs at command line level, also including temperature. However, monitoring the PCIe-SSDs with the ServerView RAID Manager is much more convenient.

## Measurement method

All the details of the measurement method and the basics of disk I/O performance are described in the white paper "[Basics of Disk I/O Performance](#)".

As standard, performance measurements of disk subsystems in PRIMERGY and PRIMEQUEST servers are carried out with a defined measurement method, which models the hard disk accesses of real application scenarios on the basis of specifications.

The essential specifications are:

- Share of random accesses / sequential accesses
- Share of read / write access types
- Block size (kB)
- Number of parallel accesses (# of outstanding I/Os)

A given value combination of these specifications is known as "load profile". The following five standard load profiles can be allocated to typical application scenarios:

Standard load profile	Access	Type of access		Block size [KB]	Application
		read	write		
File copy	random	50%	50%	64	Copying of files
File server	random	67%	33%	64	File server
Database	random	67%	33%	8	Database (data transfer) Mail server
Streaming	sequential	100%	0%	64	Database (log file), Data backup; Video streaming (partial)
Restore	sequential	0%	100%	64	Restoring of files

In order to model applications that access in parallel with a different load intensity, the "# of Outstanding I/Os" is increased, starting with 1, 3, 8 and going up to 512 (from 8 onwards in increments to the power of two).

The measurements of this document are based on these standard load profiles.

The main results of a measurement are:

- Throughput [MB/s]      Throughput in megabytes per second
- Transactions [I/O/s]      Transaction rate in I/O operations per second
- Latency [ms]              Average response time in ms

The data throughput has established itself as the normal measurement variable for sequential load profiles, whereas the measurement variable "transaction rate" is mostly used for random load profiles with their small block sizes. Data throughput and transaction rate are directly proportional to each other and can be transferred to each other according to the formula

<i>Data throughput [MB/s]</i>	$= \text{Transaction rate [I/O/s]} \times \text{Block size [MB]}$
<i>Transaction rate [I/O/s]</i>	$= \text{Data throughput [MB/s]} / \text{Block size [MB]}$

Storage media capacities are specified in this document solely with base 10 (1 TB =  $10^{12}$  bytes), whereas all other capacities, file sizes, block sizes and throughputs are specified with base 2 (1 MB/s =  $2^{20}$  bytes/s).

## Measurement environment

All the measurement results discussed in this document were determined using the hardware and software components listed below:

System Under Test (SUT)	
<b>Hardware</b>	
Model	1 x PRIMERGY RX2530 M1 1 x PRIMERGY RX2540 M1
Processor	2 x Xeon E5-2637 v3 (3.5 GHz) 2 x Xeon E5-2698 v3 (2.30 GHz) 2 x Xeon E5-2603 v3 (1.60 GHz)
Storage media	4 x PCIe-SSD 800GB P3700 1 x PCIe-SSD 1.6TB P3700 1 x PCIe-SSD 2TB P3700 1 x SSD PCIe3 800GB Main 2.5' H-P EP 1 x SSD PCIe3 1.6TB Main 2.5' H-P EP 1 x SSD PCIe3 2TB Main 2.5' H-P EP
<b>Software</b>	
BIOS	PRIMERGY RX2530 M1: R1.9.0 PRIMERGY RX2540 M1: R1.24.0
BIOS settings	For measurements of type "Performance": Intel Virtualization Technology = Disabled VT-d = Disabled Energy Performance = Performance Utilization Profile = Unbalanced CPU C6 Report = Disabled  For measurements of type "Default" (BIOS defaults): Intel Virtualization Technology = Enabled VT-d = Enabled Energy Performance = Balanced Performance Utilization Profile = Even CPU C6 Report = Enabled
Firmware	Systems: PRIMERGY RX2530 M1: iRMC S4 7.82F PRIMERGY RX2540 M1: iRMC S4 8.00d  PCIe-SSD: 8DV1FJP5
Operating system	Microsoft Windows Server 2012
Operating system settings	Power plan: For measurements of type "Performance": Select a power plan = High performance; For measurements of type "Default": Select a power plan = Balanced;
Drivers	PCIe-SSD: IaNVMe 1.1.0.1004
Administration software	Intel® Solid-State Drive Data Center Tool 2.2.2 ServerView RAID Manager 6.1.4
Initialization of RAID arrays	-
File system	NTFS
Measuring tool	Iometer 2006.07.27
Measurement data	Measurement file of 32 GB In some of the cases specified the measurement file had the full size of a partition.
Alignment of Iometer accesses	Aligned to whole multiples of 4096 bytes

Some components may not be available in all countries or sales regions.

## Measurement results

When the term “PCIe-SSDs” is used below, it refers exclusively to the models mentioned in the section entitled “[Introduction and technical data](#)”. The two form factors of the PCIe-SSD P3700 series, AIC and SFF, do not differ in their performance. Therefore, the measurement results in this white paper apply to both form factors.

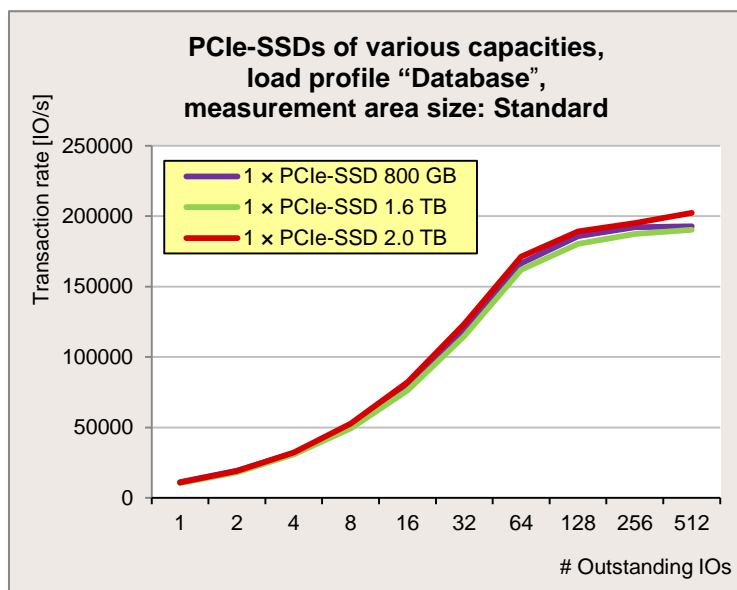
Here we use the five standard load profiles “Database”, “File server”, “File copy”, “Streaming” and “Restore”, which were described in the section “[Measurement method](#)”, to examine the performance of the PCIe-SSDs. If various load intensities are considered in the diagrams for accesses to the storage medium, they are specified, as is usual for Fujitsu, in “# of outstanding IOs”. The applications with low load intensity are represented by one outstanding IO and the applications with a very high load intensity by 512 outstanding IOs. To specifically present the performance of the data medium, optimal conditions are taken as a basis for the server. This means e.g. a CPU with maximum nominal frequency (3.5 GHz). The dependence of performance on CPU frequency will be presented at an appropriate place.

### A single PCIe-SSD

The load profiles with random access are dealt with first of all, and secondly the load profiles with sequential access.

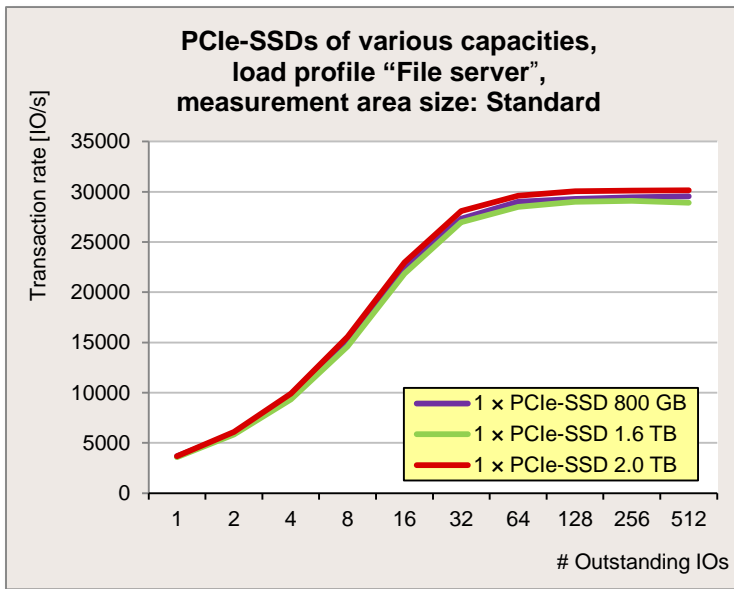
#### Random accesses

As is customary for random accesses, the transaction rate is specified below in IO/s as the measure for performance. The following diagram shows the transaction rates for the “Database” load profile (random access, 67% read, 8 kB block size). With low load intensities a PCIe-SSD provides about 10000 IO/s. As load intensity increases, the transaction rate increases evenly until it reaches just under 190000 to 200000 IO/s with a very high load intensity. In the case of this load profile the three capacity versions differ only very slightly as far as their performance is concerned.



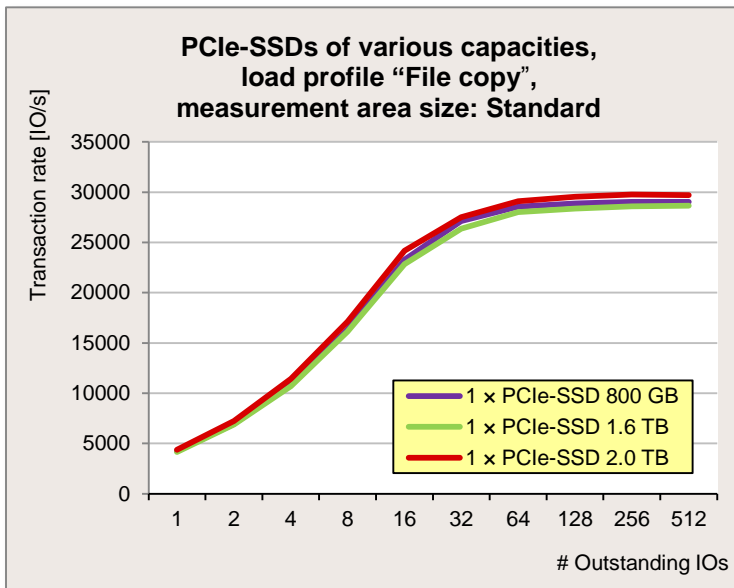
As is customary for random accesses, the transaction rate is specified below in IO/s as the measure for performance. The following diagram shows the transaction rates for the “Database” load profile (random access, 67% read, 8 kB block size). With low load intensities a PCIe-SSD provides about 10000 IO/s. As load intensity increases, the transaction rate increases evenly until it reaches just under 190000 to 200000 IO/s with a very high load intensity. In the case of this load profile the three capacity versions differ only very slightly as far as their performance is concerned.

The next diagram shows the transaction rates for the “File server” load profile (random access, 67% read, 64 kB block size). With low load intensities a PCIe-SSD provides about 3600 IO/s. As load intensity increases, the transaction rate increases evenly until it reaches about 30000 IO/s with a very high load intensity. And in the case of this load profile the three capacity versions also only differ very slightly as far as their performance is concerned.



As load intensity increases, the transaction rate increases evenly until it reaches about 30000 IO/s with a very high load intensity. And in the case of this load profile the three capacity versions also only differ very slightly as far as their performance is concerned.

The next diagram shows the transaction rates for the “File copy” load profile (random access, 50% read, 64 kB block size). With low load intensities a PCIe-SSD provides about 4200 IO/s. As load intensity increases, the transaction rate increases evenly until it reaches about 29000 IO/s with a very high load intensity. In the case of this load profile the three capacity versions also only differ very slightly as far as their performance is concerned.



As load intensity increases, the transaction rate increases evenly until it reaches about 29000 IO/s with a very high load intensity. In the case of this load profile the three capacity versions also only differ very slightly as far as their performance is concerned.

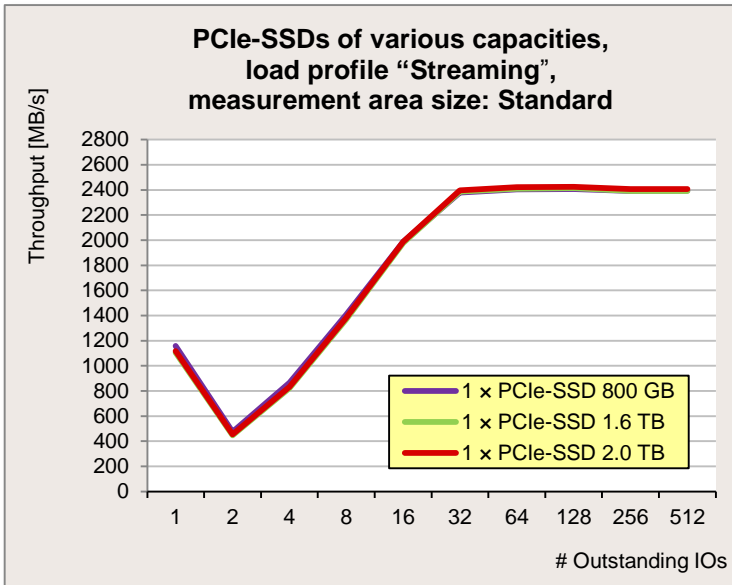


### Sequential accesses

PCIe-SSDs are scarcely used for sequential accesses in productive applications, because conventional hard disks already achieve good performance values for this purpose. Nevertheless, such load profiles are for the sake of completeness also to be discussed here.

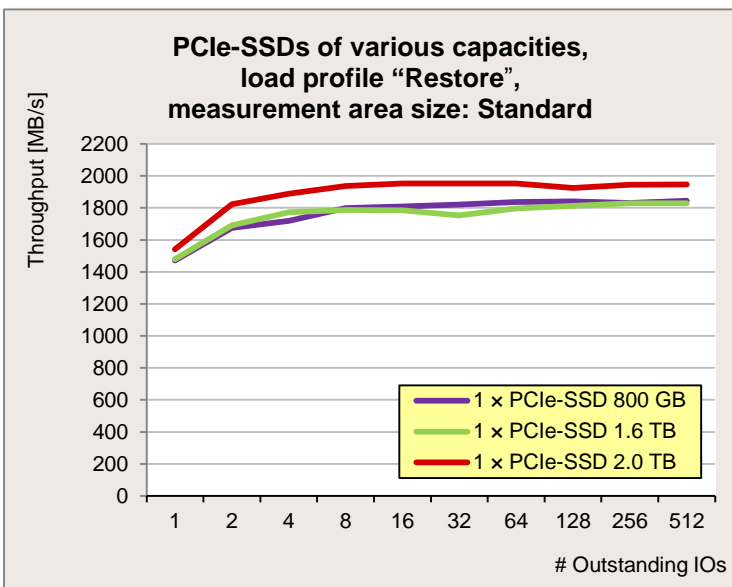
As is customary for sequential accesses, the transaction rate is no longer specified below as the measure for performance, but throughput in MB/s.

The diagram opposite shows the throughputs for the “Streaming” load profile (sequential access, 100% read, 64 kB block size). In the case of synchronous read (1 outstanding IO) a PCIe-SSD achieves a throughput of about 1100 MB/s. And with 2 outstanding IOs the throughput is about 450 MB/s. If load intensity increases to 32 outstanding IOs, the throughput increases to 2400 MB/s and also retains this value for all higher load intensities. In the case of this load profile the three capacity versions do not differ as far as their performance is concerned.



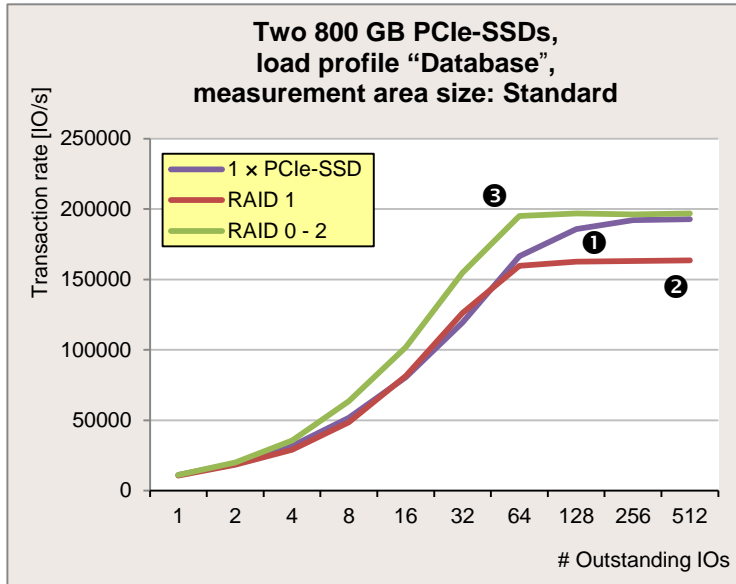
The next diagram shows the throughputs for the “Restore” load profile (sequential access, 100% write, 64 kB block size). In the case of low load intensities the two smaller capacity versions achieve a data throughput of about 1470 MB/s, which slowly increases to about 1830 MB/s as the load intensity grows.

The largest capacity version is somewhat more powerful with this load profile. It achieves a throughput of about 1540 MB/s with low load intensities, which increases to about 1950 MB/s as the load intensities grow.



## Two PCIe-SSDs

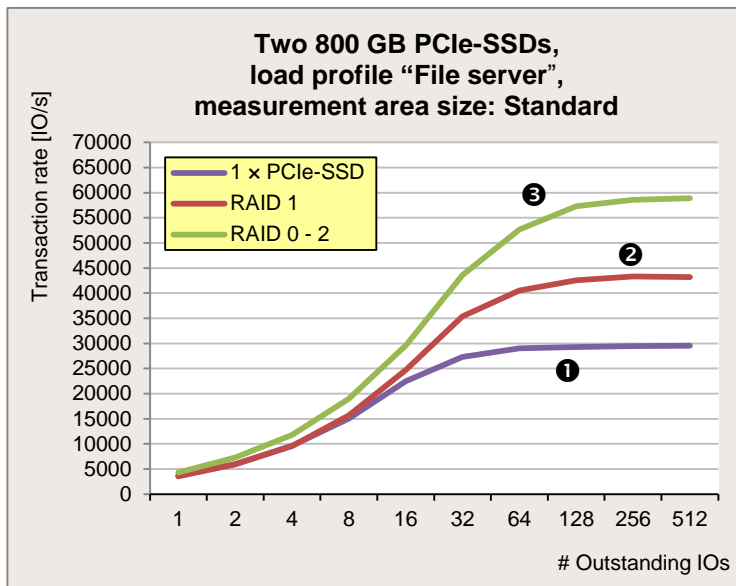
As with other logical drives, RAID arrays can also be formed from PCIe-SSDs at operating system level. To increase fail-safety you can use RAID 1. In the case of load profiles that are not exclusively write profiles this RAID level usually also results in an increase in performance for the read share of the accesses. In case of an application in which higher performance is more important than fail-safety, a RAID 0 array can also be formed from two PCIe-SSDs. The possible RAID configurations for various load intensities with a single



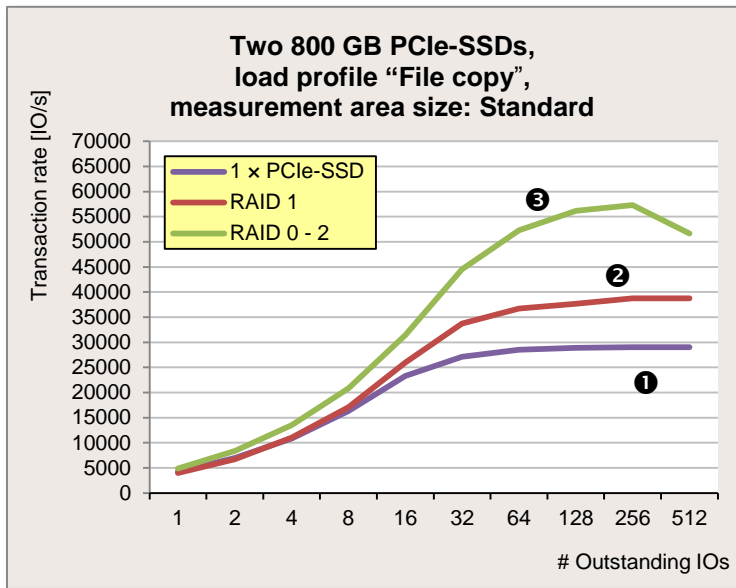
PCIe-SSD (using the example of the 800 GB version) are to be compared below for each of the five standard load profiles.

In the case of the “Database” load profile (random access, 67% read, 8 kB block size) the RAID 1 array (2) and the RAID 0 array (3) have the same transaction rate as a single PCIe-SSD (1) for the lowest load intensity, namely about 10000 IO/s. If load intensity increases up to 512 outstanding IOs, the transaction rates increasingly grow further apart: The single PCIe-SSD (1) has up to approximately 193000 IO/s, the RAID 1 array (2) achieves up to approximately 163000 IO/s, and the RAID 0 array (3) has up to approximately 197000 IO/s.

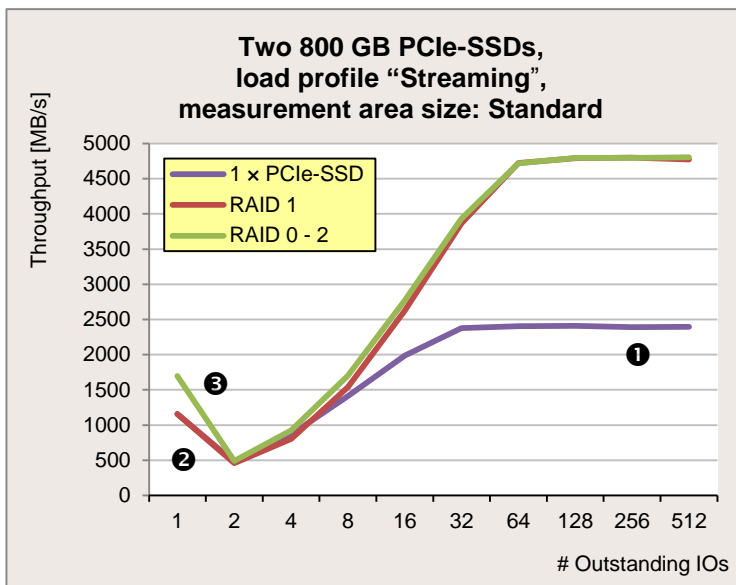
It is clear from this that if you form an operating system RAID from several PCIe-SSDs for small block sizes, the transaction rates can only be increased somewhat for medium load intensities. Consequently, the maximum transaction rate of a single PCIe-SSD cannot be appreciably exceeded.



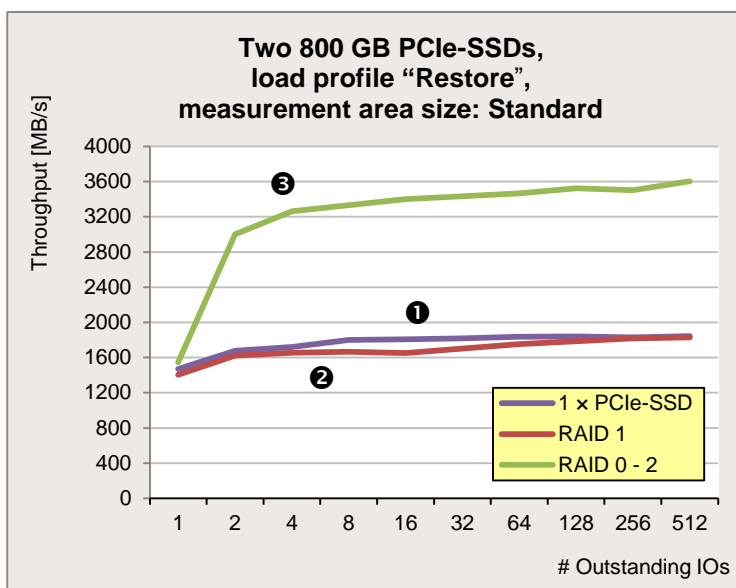
In the case of the “File server” load profile (random access, 67% read, 64 kB block size) the RAID 1 array and the single PCIe-SSD (2 and 1) have a transaction rate of about 3600 IO/s for the lowest load intensity, and the RAID 0 array (3) achieves 4300 IO/s. If you increase the load intensity, all the transaction rates increase continuously. In the case of the highest load intensity the RAID 0 array (3) achieves about 59000 IO/s, the RAID 1 array (2) about 43000 IO/s and the single PCIe-SSD (1) about 29000 IO/s. The transaction rate for RAID 1 for almost all load intensities lies between the one for a single PCIe-SSD and the one for RAID 0.



The case of the "File copy" load profile (random access, 50% read, 64 kB block size) is similar to the "File server" load profile. The RAID 1 array (②) achieves about 4000 IO/s for the lowest load intensity; the single PCIe-SSD and the RAID 0 array (① and ③) achieve 4300 IO/s and 4900 IO/s respectively. If you increase the load intensity, the transaction rates increase for the most part continuously in all three cases. In the case of the high load intensities the RAID 0 array (③) achieves about 57000 IO/s, the RAID 1 array (②) about 39000 IO/s and the single PCIe-SSD (①) about 29000 IO/s. The transaction rate for RAID 1 for almost all load intensities lies between the one for a single PCIe-SSD and the one for RAID 0.



In the case of the "Streaming" load profile (sequential access, 100% read, 64 kB block size) both the RAID arrays (② and ③) and the single PCIe-SSD (①) have about 500 MB/s (for 2 outstanding IOs) as the lowest throughput. If you increase the load intensity up to 512 outstanding IOs, RAID 0 achieves a maximum of 4800 MB/s and the single PCIe-SSD a maximum of 2400 MB/s. Up to eight outstanding IOs the throughput for RAID 1 acts in the same way as with a single PCIe-SSD; from about 16 outstanding IOs the throughput then is almost exactly the same as with RAID 0.

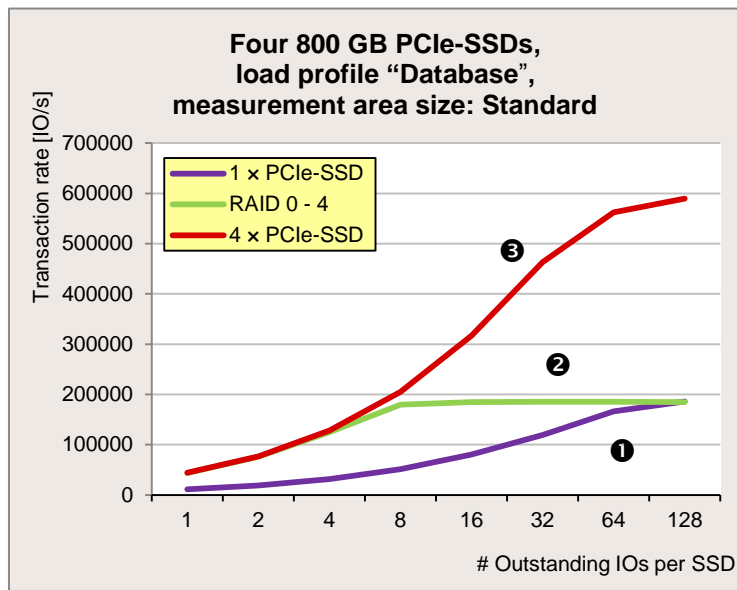


In the case of the "Restore" load profile (sequential access, 100% write, 64 kB block size) the throughputs of the three cases - single PCIe-SSD (①), RAID 0 and RAID 1 - are close together for one outstanding IO, namely between 1400 MB/s and 1550 MB/s. For all higher load intensities the throughput for RAID 0 (③) is clearly different than for the other two cases. In the case of RAID 0 3200 MB/s are already exceeded with four outstanding IOs, and the maximum lies at around 3600 MB/s. In the other two cases (① and ②) the throughput increases slightly compared with one outstanding IO; the maximum in both cases is about 1800 MB/s.

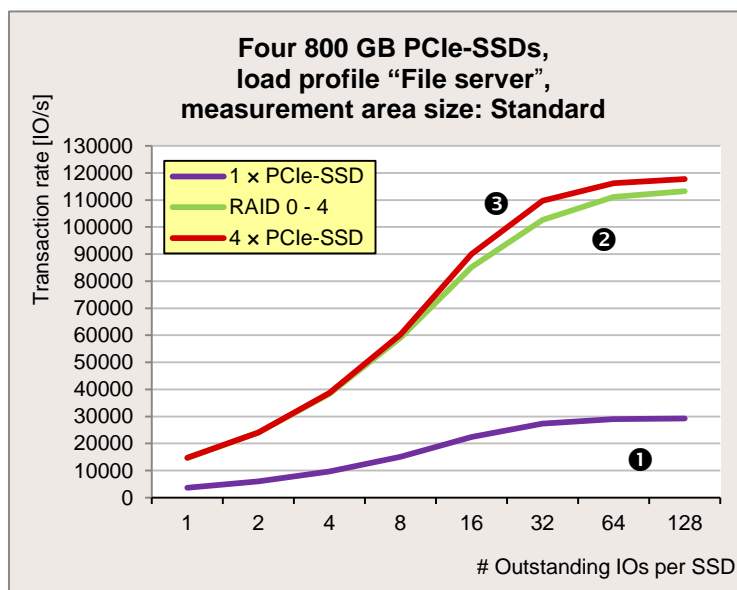
## Four PCIe-SSDs

Most of the current PRIMERGY and PRIMEQUEST servers (e. g. PRIMERGY RX2540 M1) can be equipped with four or more PCIe-SSDs. A RAID array is not always created from all the PCIe-SSDs in the case of such a number. You often have several processes or applications that create disk I/Os and which you prefer to have run on clearly separated PCIe-SSDs. It is interesting to know for such cases whether the available disk I/O performance scales with the number of PCIe-SSDs in the server. As an example the following five diagrams contain performance comparisons between a single PCIe-SSD and four PCIe-SSDs. The four PCIe-SSDs are on the one hand considered as four single media under independent load, and on the other hand as a RAID 0 array via an operating system RAID. To enable a fair comparison of the latter two cases the same load intensity is in each case compared for each PCIe-SSD involved (as shown in the diagrams of this section by the caption “# Outstanding IOs per SSD” on the right axis).

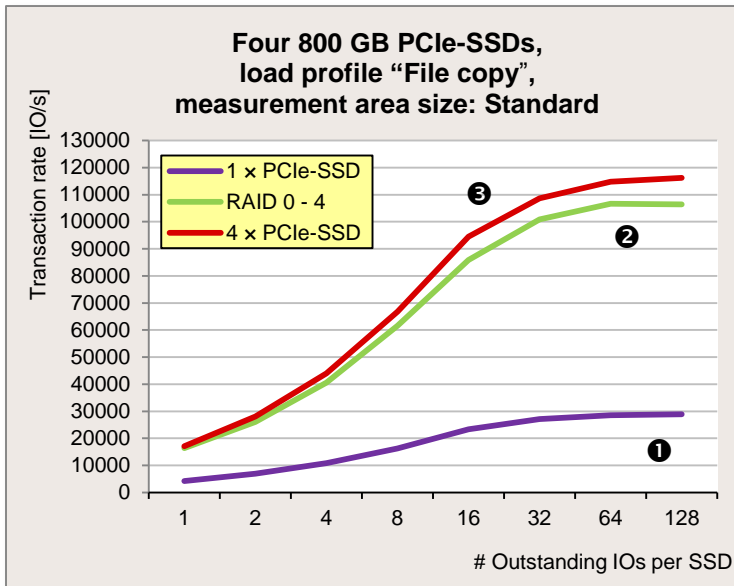
In the case of the “Database” load profile (random access, 67% read, 8 kB block size) the single PCIe-SSD (❶) has a transaction rate of about 11000 IO/s for the lowest load intensity, and a transaction rate of about 186000 IO/s for the largest load intensity.



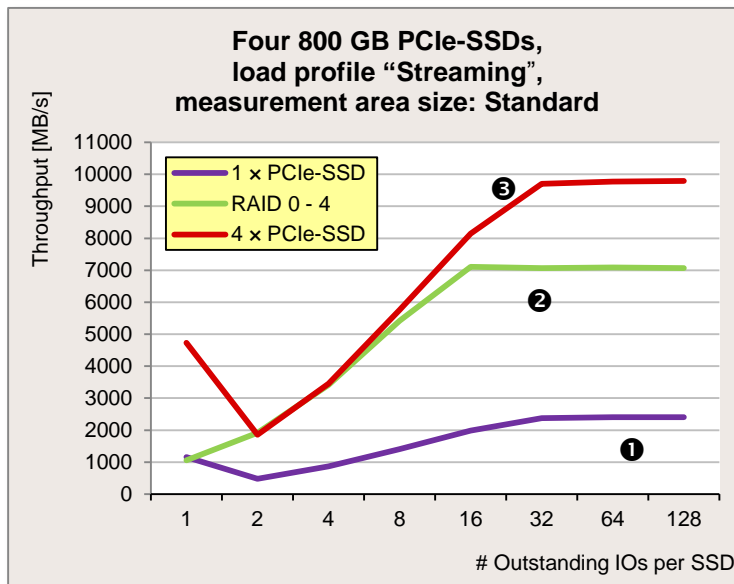
186000 IO/s for the largest load intensity. In the case of four PCIe-SSDs (❸) under independent load the corresponding transaction rates are 44600 IO/s (scaling factor 4.0) and 589000 IO/s (scaling factor 3.17). The operating system RAID consisting of four PCIe-SSDs (❷) has for the lowest load intensity approximately the same transaction rate as the four PCIe-SSDs under independent load (❸), i.e. 43800 IO/s. The transaction rate is approximately as large for the highest load intensity as with a single PCIe-SSD (❶), i.e. 185000 IO/s. Thus, the operating system RAID consisting of four PCIe-SSDs can (just as in the case of two PCIe-SSDs) not exceed the maximum transaction rate of a single PCIe-SSD (❶) for this small block size.



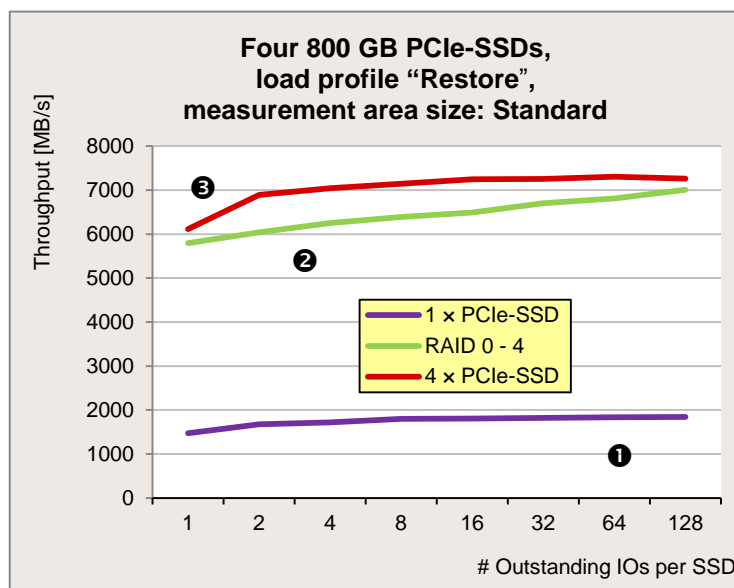
In the case of the “File server” load profile (random access, 67% read, 64 kB block size) the four PCIe-SSDs under independent load (❸) provide a transaction rate of about 14700 IO/s (scaling factor 3.99) for the lowest load intensity and about 118000 IO/s (scaling factor 4.02) for the highest load intensity. The RAID array consisting of four PCIe-SSDs (❷) provides 14600 IO/s for low load intensities and increases to 113000 IO/s for the highest load intensity, thus coming quite close to the four PCIe-SSDs under independent load.



In the case of the "File copy" load profile (random access, 50% read, 64 kB block size) the four PCIe-SSDs under independent load (③) provide a transaction rate of about 17000 IO/s (scaling factor 4.00) for the lowest load intensity and about 116000 IO/s (scaling factor 4.02) for the highest load intensity. The RAID array consisting of four PCIe-SSDs (②) provides about 16400 IO/s for low load intensities and increases to 106000 IO/s for the highest load intensity, thus coming - similar to the case of the "File server" - close to the four PCIe-SSDs under independent load.



In the case of the "Streaming" load profile (sequential access, 100% read, 64 kB block size) the four PCIe-SSDs under independent load (③) once again show an almost perfect scaling factor. For read with 1 outstanding IO per SSD the four PCIe-SSDs achieve a data throughput of about 4700 MB/s (scaling factor 4.08). The data throughput of the four PCIe-SSDs with 2 outstanding IOs per SSD is about 1860 MB/s (scaling factor 3.90). If you increase the load intensity to 32 outstanding IOs per SSD, data throughput increases to 9700 MB/s and also approximately retains this value for all higher load intensities (scaling factor 4.07 with 128 outstanding IOs per SSD). The throughput curve for the operating system RAID (②) lies somewhere in between and has a maximum of 7100 MB/s.



In the case of the "Restore" load profile (sequential access, 100% write, 64 kB block size) the four PCIe-SSDs under independent load (③) achieve about 6100 MB/s with 1 outstanding IO per SSD and increase to 7264 MB/s with 128 outstanding IOs per SSD. The scaling factor is again almost always around 4.0. The throughput curve for the operating system RAID (②) is once again between the two other curves and achieves about 7000 MB/s as a maximum.

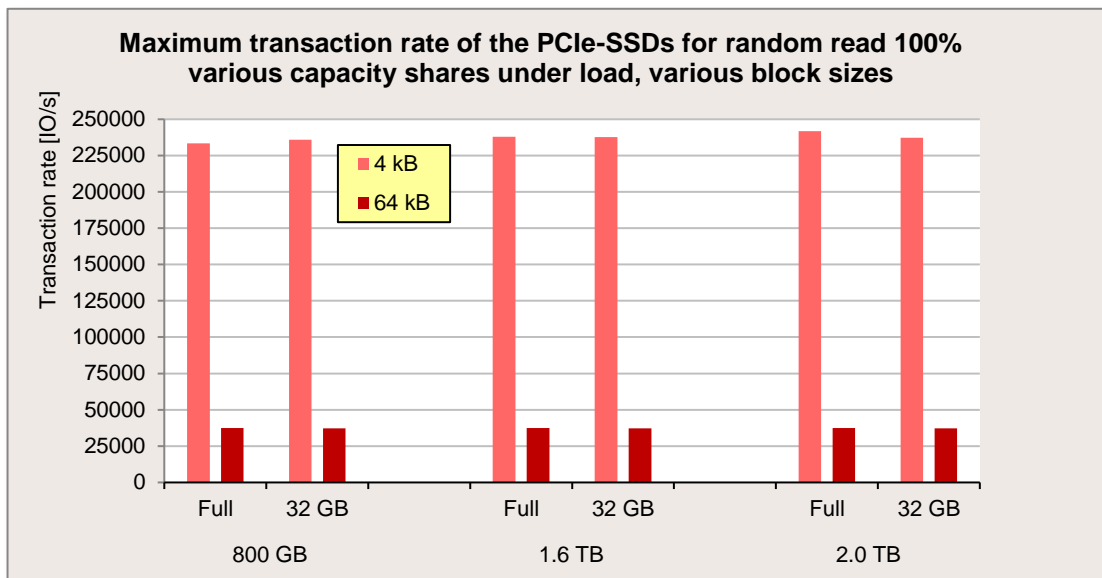
## Storage capacity share under load

In most cases the majority of the accesses to a PCIe-SSD in a given period of time should concentrate on a relatively small subarea of the available storage addresses („hot spot”). Example: In a mail-order company for clothing the majority of the accesses to the order processing database refers to seasonal fashion articles. This concentration of accesses is modelled by the small size of the measurement file subject to the measurement methodology used (see section “[Measurement method](#)”) and was consequently also the basic assumption in the previous sections.

However, if all the available storage capacity of an SSD is equally under load with accesses, this is the worst possible case from a performance point of view. This case is normally dealt with in the data sheets. The significance of the influencing variable “Storage capacity share under load” for the performance of a PCIe-SSD is to be illustrated in the following three diagrams for various load profiles.

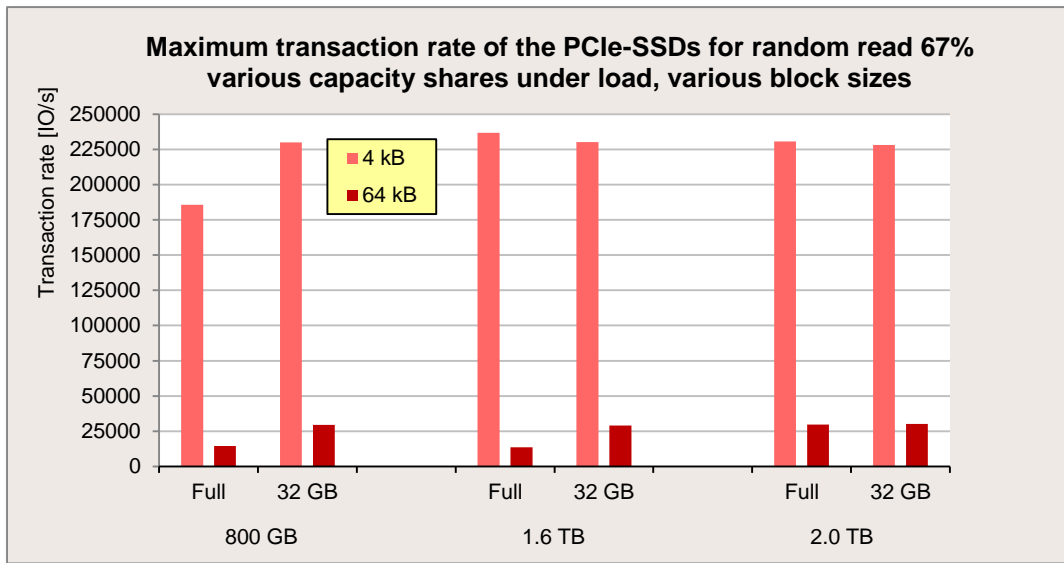
In the case of sequential load profiles there are no appreciable differences in performance between the various sizes of the storage capacity share under load. The following three diagrams are therefore restricted to comparisons for random load profiles with various write shares. In each case, the diagrams compare a PCIe-SSD that is under an even load for the entire storage capacity and, according to the measurement method, a PCIe-SSD of which a small subarea of 32 GB is under load.

The first diagram considers the case “random access, 100% read”. In this case, there are no appreciable differences in performance between the various sizes of the storage capacity share under load for all three storage capacities and all block sizes.

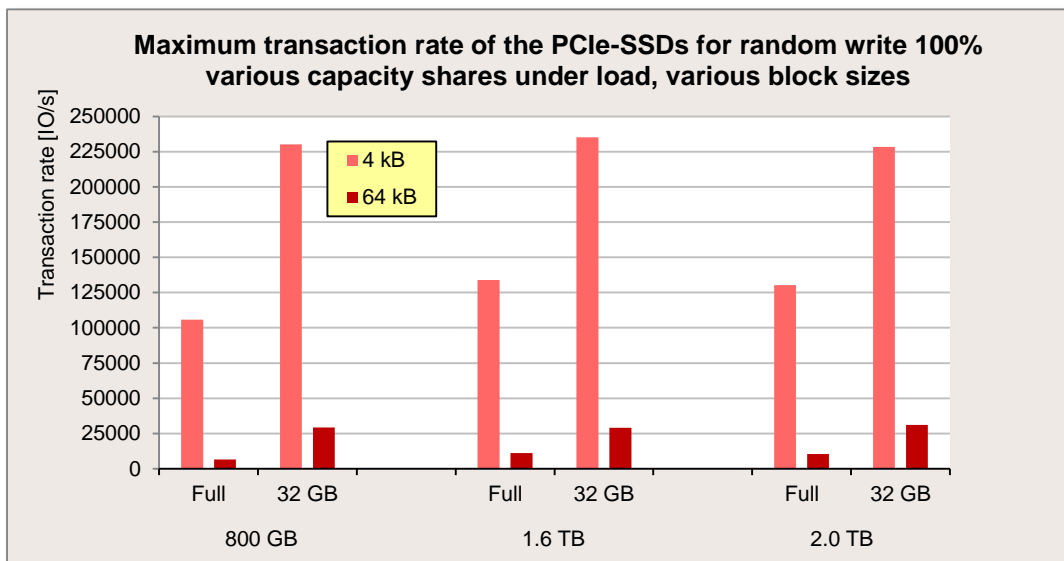




The next diagram considers the case “random access, 67% read”. For the 800 GB version here you can see significant differences between the various sizes of the subarea under load for all block sizes. Notable differences between the various sizes of the area under load can only be seen for large blocks (64 kB in this example here) for the 1.6 TB version.



The last diagram considers the case “random access, 100% write”. Significant differences between the various sizes of the area under load can be seen here for all capacity versions.



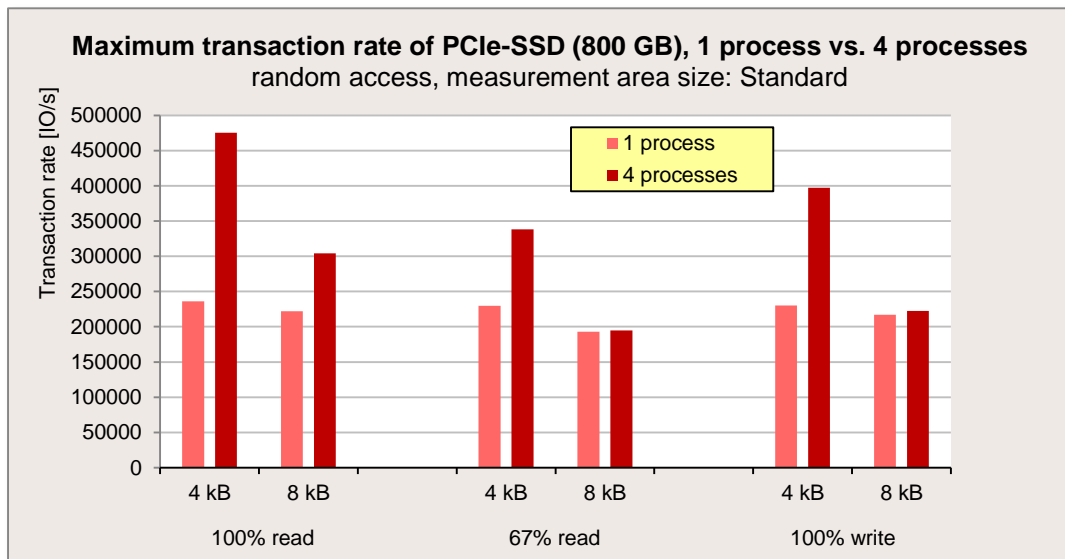
To sum up:

The PCIe-SSDs show the highest transaction rates for random accesses with a write share if the SSD subarea under load is small. The performance disadvantage of PCIe-SSDs under an equal load over the entire storage area is greater the higher the write share, and the larger the block size of the accesses. This disadvantage is somewhat less pronounced for the larger capacity versions.

## Access of several processes to a PCIe-SSD

The section entitled “[Storage capacity share under load](#)” illustrated the worst case performance level of a PCIe-SSD, i.e. with an even load of the entire storage capacity by a single process. In contrast, the standard measurement method that is otherwise used in this document models access concentration onto a small subarea and consequently provides significantly better performance values compared with the case of an “even load”. These performance values which are already very good can be significantly exceeded again for small block sizes if - in addition to concentrating the load on a small percentage of the storage capacity - you have accesses from various server processes (e. g. several applications) to separate subareas of the PCIe-SSD.

The example used below models an application scenario, in which several applications or application instances (generally referred to in the diagram as “processes”) access dedicated, separate partitions of a PCIe-SSD. The standard measurement method used also models access concentration onto a small subarea of 32 GB in each partition. The accesses of the applications are random with a small block size. Furthermore, a high-frequency CPU (in this case, 3.5 GHz nominal frequency) is necessary to achieve the transaction rates shown in the following diagram. The example is dramatic and shows transaction rates that are difficult to achieve in practice. The reason for this is that the privileged CPU time required for the many I/Os alone accounts for a significant share of the overall server-CPU performance. Nevertheless, the example shows the principal options of the PCIe-SSDs under consideration. The depicted transaction rates are the maximums of all the numbers of outstanding IOs for the respective access pattern.



The example shows that considerable increases in performance are possible by using four processes instead of one process for random accesses with small blocks. Increases of up to about 100% for 4 kB (most in evidence with 100% read); and still around 37% for 8 kB and 100% read.

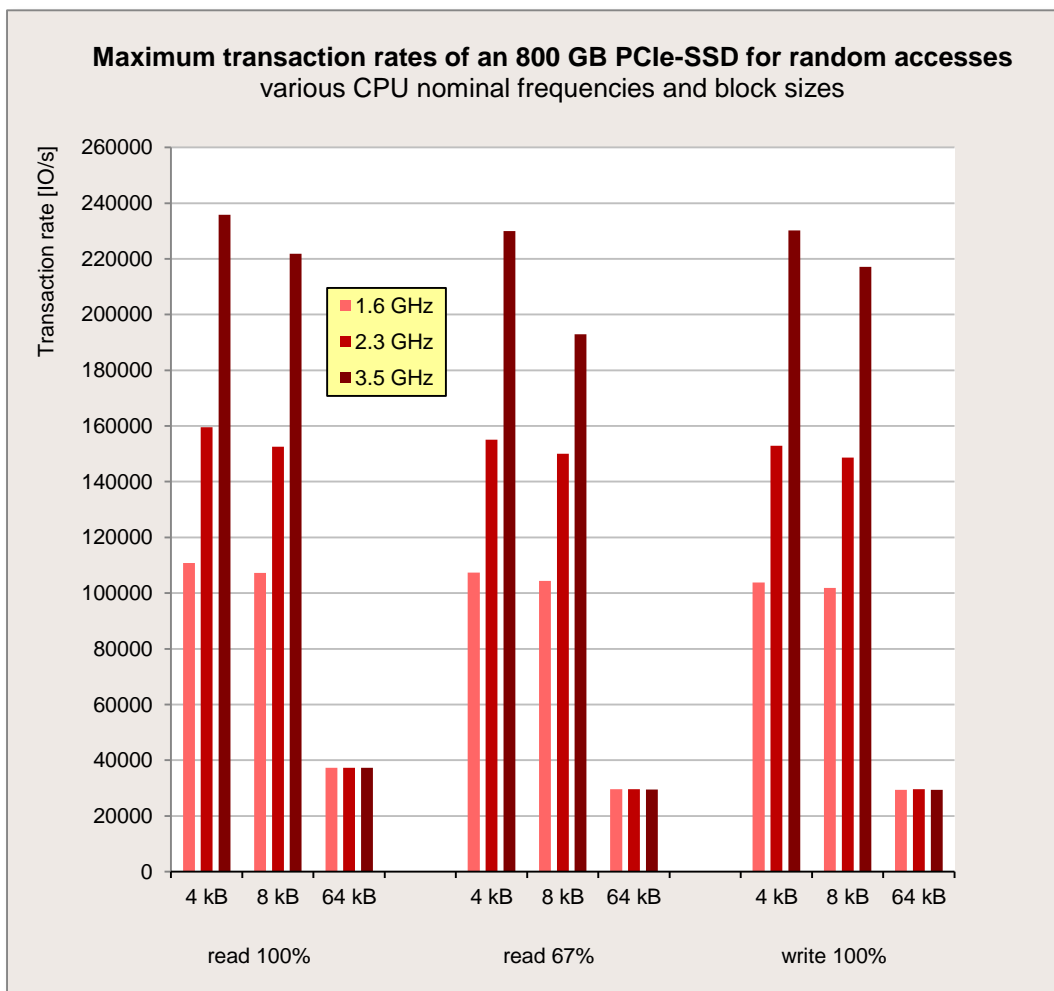


## Best practice

Like every I/O component of a server, PCIe-SSDs also depend on the hardware conditions of the server. The processor, main memory and PCIe bus in particular play a role here. The performance of PCIe-SSDs can be positively influenced by the right selection, configuration and setting of the server.

### Optimizing the current processor frequency

The actual processor frequency has an influence on the performance of PCIe-SSDs. In the ideal case, the PRIMERGY or PRIMEQUEST system is equipped with the processor type with the highest possible nominal frequency. However, this nominal frequency is in itself not yet significant, because modern processors can both temporarily increase and decrease the actual frequency compared with the nominal frequency by using technologies such as Turbo mode and energy-saving functions. The reduction can e. g. occur in applications with smaller computing requirements. If performance is more important than energy saving in an application, this lowering of frequency can be prevented by changing the settings. The following examples illustrate the performance effect you achieve through optimizing the CPU frequency. The first issue to be examined is the impact of choosing a processor with higher nominal frequency. For this purpose, the following diagram considers a PCIe-SSD 800 GB in a current dual-socket server. The server is equipped with two CPUs of the Xeon E5-2600 v3 family, which by means of suitable BIOS and operating system settings (listed in the section "[Measurement environment](#)") is tuned to maximum performance. The diagram compares the transaction rates for CPUs with 1.6 GHz, 2.3 GHz and 3.5 GHz for selected load profiles.

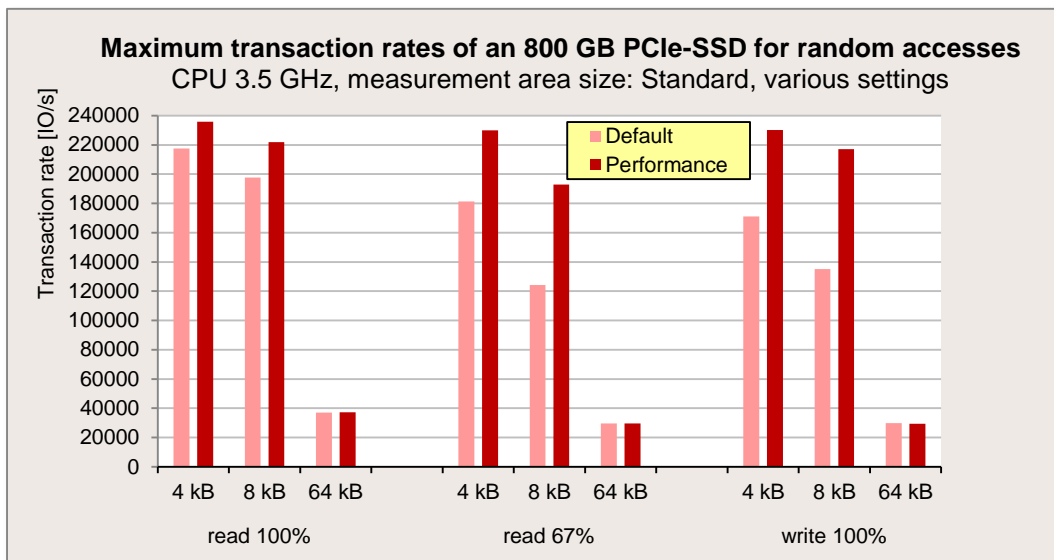


The diagram demonstrates that a significant advantage can be gained for block sizes up to 8 kB through a higher nominal frequency. The nominal CPU frequency has no impact for load profiles with larger blocks.

It should also be mentioned that advantages can only be generated for transaction rates by means of a high CPU frequency if the desired IO/s value can be achieved at all on account of other marginal conditions. For example, the aspect mentioned in the section "[Storage capacity share under load](#)" is of significance.

The second issue to be examined for a given CPU is the BIOS and operating system setting options and their impact on the current CPU frequency. More information about these setting options for performance or energy savings is available in the white paper [“BIOS optimizations for Xeon E5-2600 v3 based systems”](#). The BIOS and operating system settings for the measurements in the previous diagram were all set to performance. The set of settings, which results in this effect, is referred to here collectively with “Performance” and is described in the section [“Measurement environment”](#). Since this set of settings is associated with a higher power consumption, the question arises as to performance with default BIOS and operating system settings (this set of settings is referred to here as “Default”).

The next diagram answers this question by comparing the maximum transaction rates (i.e. the maximums over all # outstanding IOs) of the 800 GB PCIe-SSD for the two mentioned sets of settings.



The diagram shows that the “Performance” set of settings enables higher transaction rates than the “Default” set of settings for block sizes up to 8 kB with load profiles with a significant write share. Prerequisite for this is that transaction rates of more than 120000 IO/s are as a matter of principle possible as far as the other marginal conditions are concerned (access concentration). To achieve high values it suffices to select the Windows “High Performance” power plan when working with BIOS defaults.

In both diagrams in this section random accesses are merely used as examples; the statements also apply for sequential accesses. Furthermore, it was for both diagrams essential to create the optimal prerequisites required to achieve a high transaction rate when further elaborating CPU-related differences in the measurement area on the PCIe-SSD. Therefore, both diagrams are again based on the “Standard” measurement area size so that the I/O accesses are concentrated on a small subarea of the storage capacity.

## Comparison with other storage media

The PCIe-SSD P3700 is to be compared below with its own predecessor as well as with one currently available HDD and SSD with an SAS-12G interface.

The following table compares the performance values of the four storage media for the five standard load profiles that have already been used.

Maximum performance for 1 – 512 outstanding IOs							
Load profile	SAS-12G-HDD 450 GB, 15 krpm, 2.5"	SAS-12G-SSD MLC 400 GB 2.5"	PCIe-SSD ioDrive <sup>®</sup> 2 1.2 TB	PCIe-SSD P3700 800 GB	Quotient PCIe-SSD P3700 / SAS-12G- HDD	Quotient PCIe-SSD P3700 / SAS-12G- SSD	Quotient PCIe-SSD P3700 / PCIe-SSD ioDrive <sup>®</sup> 2
	Disk cache enabled	Disk cache enabled	Formatting "Factory Capacity"	1 Process, Measurement area 32 GB			
Database	864 IO/s	64416 IO/s	49919 IO/s	192881 IO/s	223.2	3.0	3.9
File server	766 IO/s	8977 IO/s	19005 IO/s	29530 IO/s	38.6	3.3	1.6
File copy	789 IO/s	8157 IO/s	18018 IO/s	29033 IO/s	36.8	3.6	1.6
Streaming	237 MB/s	790 MB/s	1390 MB/s	2395 MB/s	10.1	3.0	1.7
Restore	237 MB/s	419 MB/s	1037 MB/s	1844 MB/s	7.8	4.4	1.8

As shown in the table, the performance values of the PCIe-SSD – depending on the load profile – achieve approximately 3.9 times that of the predecessor generation, ioDrive<sup>®</sup>2, almost the 4.4 times that of the SAS-12G-SSD and more than 220 times that of the SAS-12G-HDD.

The following table compares the minimum latency times when reading and writing small blocks.

Load profile	SAS-12G-HDD 450 GB, 15 krpm, 2.5"	SAS-12G-SSD MLC 400 GB 2.5"	PCIe-SSD ioDrive <sup>®</sup> 2 1.2 TB	PCIe-SSD P3700 800 GB
	Disk cache enabled	Disk cache enabled	Formatting "Factory Capacity"	1 Process, Measurement area 32 GB
Read Latency (1 kB sequential)	0.105 ms	0.21 ms	0.066 ms	0.041 ms
Write Latency (1 kB sequential)	0.51 ms	0.11 ms	0.018 ms	0.016 ms

Thus, the minimum latency was further reduced in comparison with the predecessor ioDrive<sup>®</sup>2.

## Conclusion

The PCIe-SSDs P3700 are an advanced non-volatile storage medium, which offers very high performance in a small space, above all for I/O accesses with a large number of transactions. Another innovation is its availability as an SFF form factor. Compared with the predecessor generation, performance has with regard to both the maximum transaction rate and maximum data throughput been significantly increased. When working with one application instance (process) and accesses concentrated on one small part of the medium, a PCIe-SSD P3700 achieves e. g. transaction rates of up to about 200000 IO/s for accesses that are typical of databases, and data throughputs of up to 2490 MB/s for sequential read accesses - as with video streaming.

The transaction rate that can be achieved in individual cases with small block sizes greatly depends on the size of the area on the PCIe-SSD, upon which the accesses are concentrated. Furthermore, the current CPU frequency of the server also plays a role when it comes to the highest transaction rates.

Considerable increases in the total performance of a PCIe-SSD are also possible if various processes access separate subareas of the medium. For example, if four processes create random accesses with small blocks on separate subareas, a total transaction rate of up to 475000 IO/s is possible.

The disk I/O performance per server can be further increased by forming RAID arrays with operating system means or by operating these storage media independently. It is possible to operate eight of these PCIe-SSDs in a number of PRIMERGY models.

## Literature


### PRIMERGY & PRIMEQUEST Servers

<http://www.fujitsu.com/fts/products/computing/servers/>


### Performance of Server Components

<http://www.fujitsu.com/fts/products/computing/servers/mission-critical/benchmarks/x86-components.html>

This White Paper:

 <http://docs.ts.fujitsu.com/dl.aspx?id=e123e7ef-20e4-4a40-9083-876a2c106494>

 <http://docs.ts.fujitsu.com/dl.aspx?id=daabec11-a857-4781-af72-64b7d86a194c>

 <http://docs.ts.fujitsu.com/dl.aspx?id=5c4d91e4-3fa0-49f9-a92e-6d51dbcdb7bd>

### PCIe-SSD P3700 Series

Datenblatt (EN)

<http://docs.ts.fujitsu.com/dl.aspx?id=b332a27a-96a5-4dfc-8af2-5fc30de74226>:

### Solid State Drives - FAQ

<http://docs.ts.fujitsu.com/dl.aspx?id=78858d6c-4c0f-479a-8ceb-705fe1938f4e>

### BIOS optimizations for Xeon E5-2600 v3 based system

<http://docs.ts.fujitsu.com/dl.aspx?id=f154aca6-d799-487c-8411-e5b4e558c88b>

### Basics of Disk I/O Performance

<http://docs.ts.fujitsu.com/dl.aspx?id=65781a00-556f-4a98-90a7-7022feacc602>

### Fusion-io<sup>®</sup> ioDrive<sup>®</sup> 2 Solid-State Storage devices (predecessor generation)

Datenblatt (EN)

<http://docs.ts.fujitsu.com/dl.aspx?id=cb40337c-4292-44fe-ac8f-50a96bc653e1>

Performance Report PCIe-SSDs ioDrive<sup>®</sup> 2

<http://docs.ts.fujitsu.com/dl.aspx?id=2d717c91-8da2-4201-8329-68823ada6ec3>

### Information about Iometer

<http://www.iometer.org/>

## Contact

### FUJITSU

Website: <http://www.fujitsu.com/>

### PRIMERGY & PRIMEQUEST Product Marketing

<mailto:Primergy-PM@ts.fujitsu.com>

### PRIMERGY Performance and Benchmarks

<mailto:primergy.benchmark@ts.fujitsu.com>

ioDrive2 is a registered trademark of SanDisk.