Tiny Drive Technology

* * *

February 19, 1993

Prepared by
Philip E. White
ECC Technologies, Inc.
4750 Coventry Road East
Minnetonka, MN 55345-3909

1. Introduction

This document describes advanced technology for tiny drives. In this document the term “tiny drives” refers to magnetic recording devices which write information to and read information from recording objects which are in motion relative to the read/write (R/W) head or heads. Information is recorded on the surfaces of the recording objects.

Special emphasis is given to a revolutionary new class of drives that ECC Technologies refers to as tiny-dumb drives. ECC Technologies proposes a tiny-dumb drive that is specially designed and optimized for future RAID systems. RAID systems designed with tiny-dumb drives as memory components will have capacities per unit of volume much greater than existing RAID systems, higher performance as measured in number of I/Os per second and a variable level of fault-tolerance.

Many financial analysts in the disk industry (including Montgomery Securities and IDC) have estimated total worldwide sales of disk array products to be nearly $8 billion in 1994 and nearly $20 billion in 1996. These facts justify our interest in drives for drive arrays.

Some of the most significant characteristics of the tiny-dumb drives being proposed by ECC Technologies are listed below:

Tiny-Dumb Drives...

- Are tiny and dumb
- Contain no address fields
- Contain no error correction fields
- Contain no error correction circuits
- Contain very short PLL synchronization fields
- Have a dumb interface
- Use a “rod” as a recording object instead of a “disk”
- Have a single R/W data head
- Have a read only (RO) clock head
- Have a fast sync phase lock loop
- Integrate the motor assembly with the rod and bearings assembly
- Use the RO clock signal for motor position feedback
- Use the RO clock signal (phase-adjusted) as the write clock
- Use the RO clock signal (phase-adjusted) as the read clock
- Use the RO clock signal to help locate the position of the R/W head
- Use a context-dependent postcompensation channel equalizer

- Are designed to be able to trade off density (capacity) for error rate
- Spin at very high RPM to reduce rotational latency
- Have an interface similar to a DRAM interface
- Are considered to be memory “components” for arrays
- If successful, will eventually be manufactured at a rate of hundreds of millions of units per year
- Will become the low-cost leader when economies of scale kick-in
- Will probably become disposable rather than repairable
- When used with parallel, Reed-Solomon ECC, need not be very reliable
- When used with parallel, Reed-Solomon ECC, may have very high raw error rates
- Do not require the use of (d,k) run-length limiting (RLL) codes
- Will probably use d-constrained RLL sequences only
- Ganged together in groups with RS ECC, will replace existing disk drives
- Can be ganged together with any number per drive group
- Will consume approximately the same space as a large IC

Different types of recording objects have been examined as possible candidates for use in tiny drives. “Rods” are found to be much more efficient than “disks” as geometries are scaled down. Other characteristics of rods make them the ideal choice for future miniature recording devices.

Rod cartridges are proposed for a new type of removable storage product known as a parallel drive storage unit (PDSU). The cartridges are failure and fault-tolerant.

This document focuses on rod devices, but the reader should keep in mind that most of the ideas and concepts described in this document can be applied to disks and other types of recording objects as well.

If companies in the disk industry who are not vertically integrated decide to implement drives based on the technology described in this document, they will probably need to collaborate in making changes to the disk drive industry infrastructure that supplies heads, media and motors. This type of activity would be well-rewarded because, if successful in the marketplace, common “drive memory components” would certainly be manufactured at the rate of hundreds of millions of units per year when in full production. Economies of scale would drive down the cost of such devices to an unprecedented level.

We begin by examining characteristics of drive arrays since we are especially interested in drive technology most appropriate for drive arrays. The “drives” used in these arrays are illustrated as disk drives but may equally well be some other type of drives such as rod drives.

2. RAID Systems

Without fault-tolerance, drive arrays with a large number of drives may be too unreliable to be practical. In order to achieve fault-tolerance, drive arrays are divided into drive groups, and drive groups are made fault-tolerant by using some type of error correction system. An illustration of a disk array divided into drive groups is shown in Figure 1.

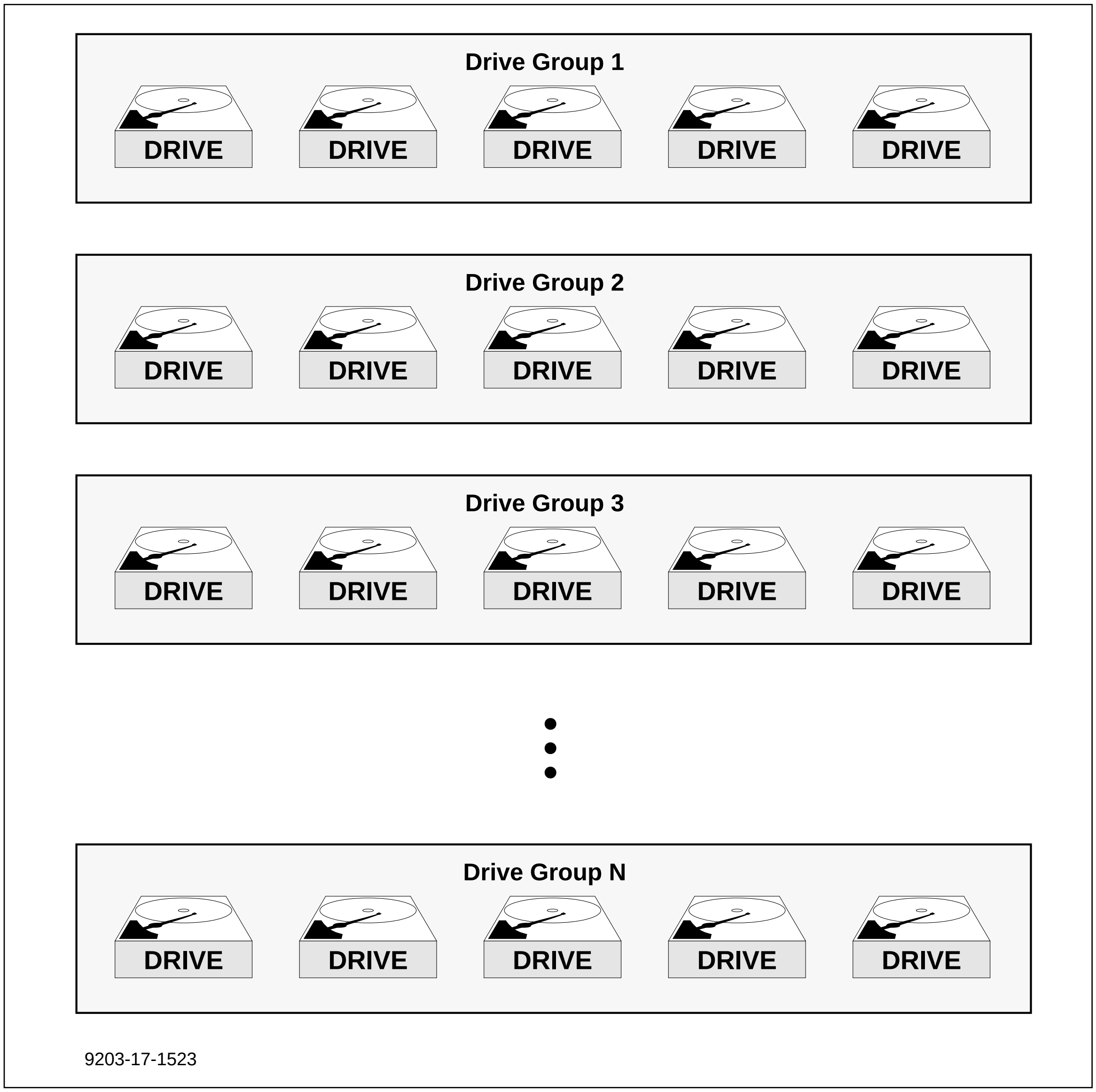


Figure 1 Disk Array Divided into Drive Groups

There are really only two types of drive groups. In one type of drive group, data is written to and read from all of the drives in the group concurrently or simultaneously and all of the data in a set of sectors is considered to be one record. This type of parallel transfer drive group is referred to as a type P drive group where “P” stands for parallel. In the other type of drive group, the drives are viewed as stand-alone devices and are accessed independently. This type of drive group is referred to as a type S drive group where “S” stands for serial. These two types of drive groups are illustrated in Figure 2.

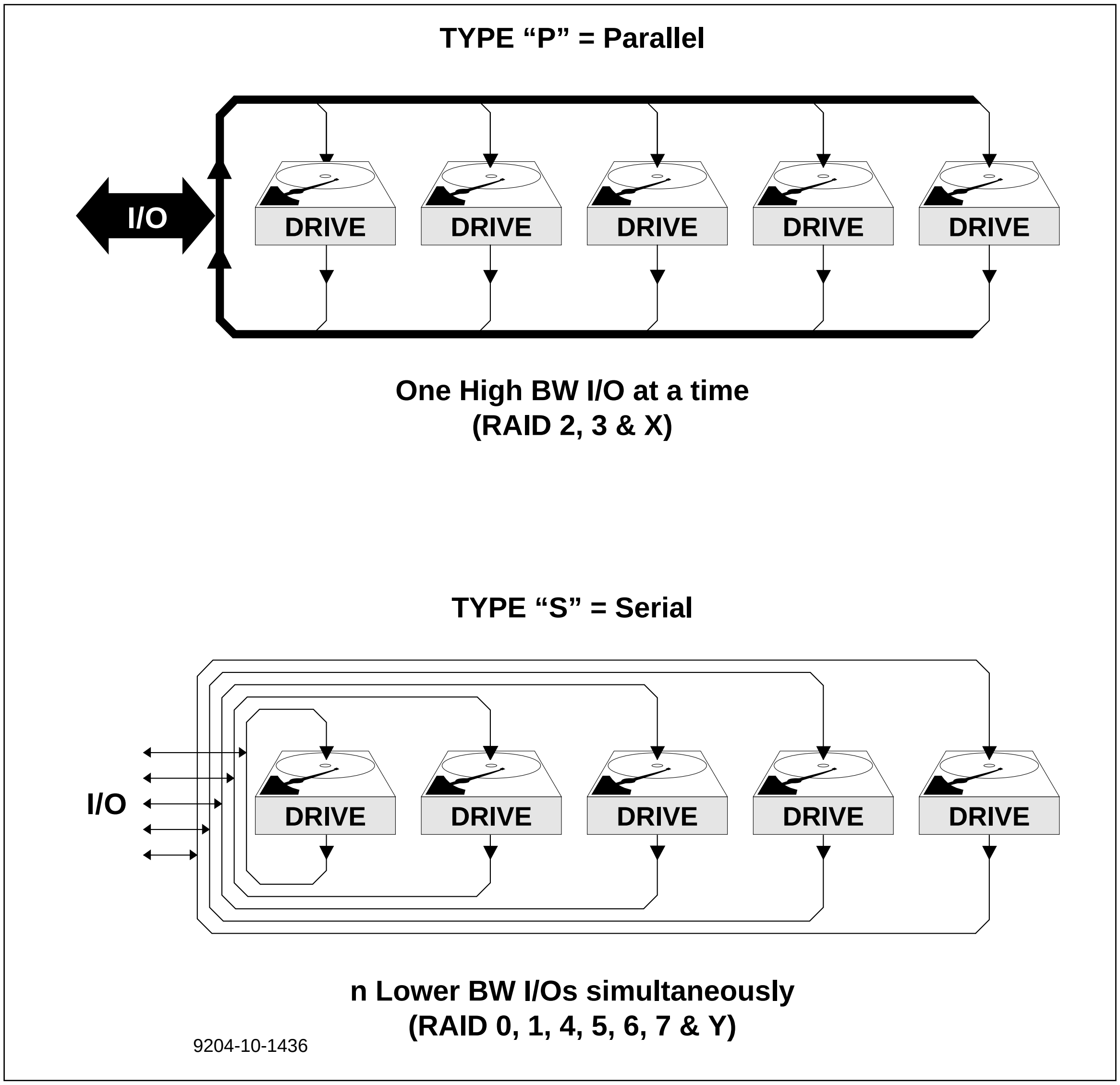


Figure 2 Two Types of Drive Groups

The illustrations in Figure 2 illustrate concepts only. They are not meant to show the way actual disk drives are interconnected in commercially-available disk arrays.

The actual physical realization of a type S drive group can be accomplished in several different ways. All of the drives in a type S drive group may be connected to a single controller via a common bus (usually SCSI) or each drive in a type S drive group may have its own controller. When only one controller is used with a type S drive group, intelligent disk drives with full-track or near full-track buffers are usually assumed so that the controller can make data transfers to/from a drive, disconnect and quickly reconnect to another drive for the next data transfer. In that way, the bandwidth (BW) of the bus does not significantly impact performance since drives spend most of their time accessing records.

Normally all the drives in a type P drive group require their own separate controller, but if there are multiple groups in an array, each column of drives may share the same controller.

ECC Technologies refers to disk arrays that use all type P drive groups as type P arrays and disk arrays that use all type S drive groups as type S arrays. In this document we compare type P arrays with N drive groups to type S arrays with N drives. (We explain why this is a “fair” comparison in Section 3.) A type P array with 20 drive groups is illustrated in Figure 3 and a type S array with 20 drives is illustrated in Figure 4. (For more information on type P and type S arrays, see ECC Technologies’ RAID & ECC Seminar material).

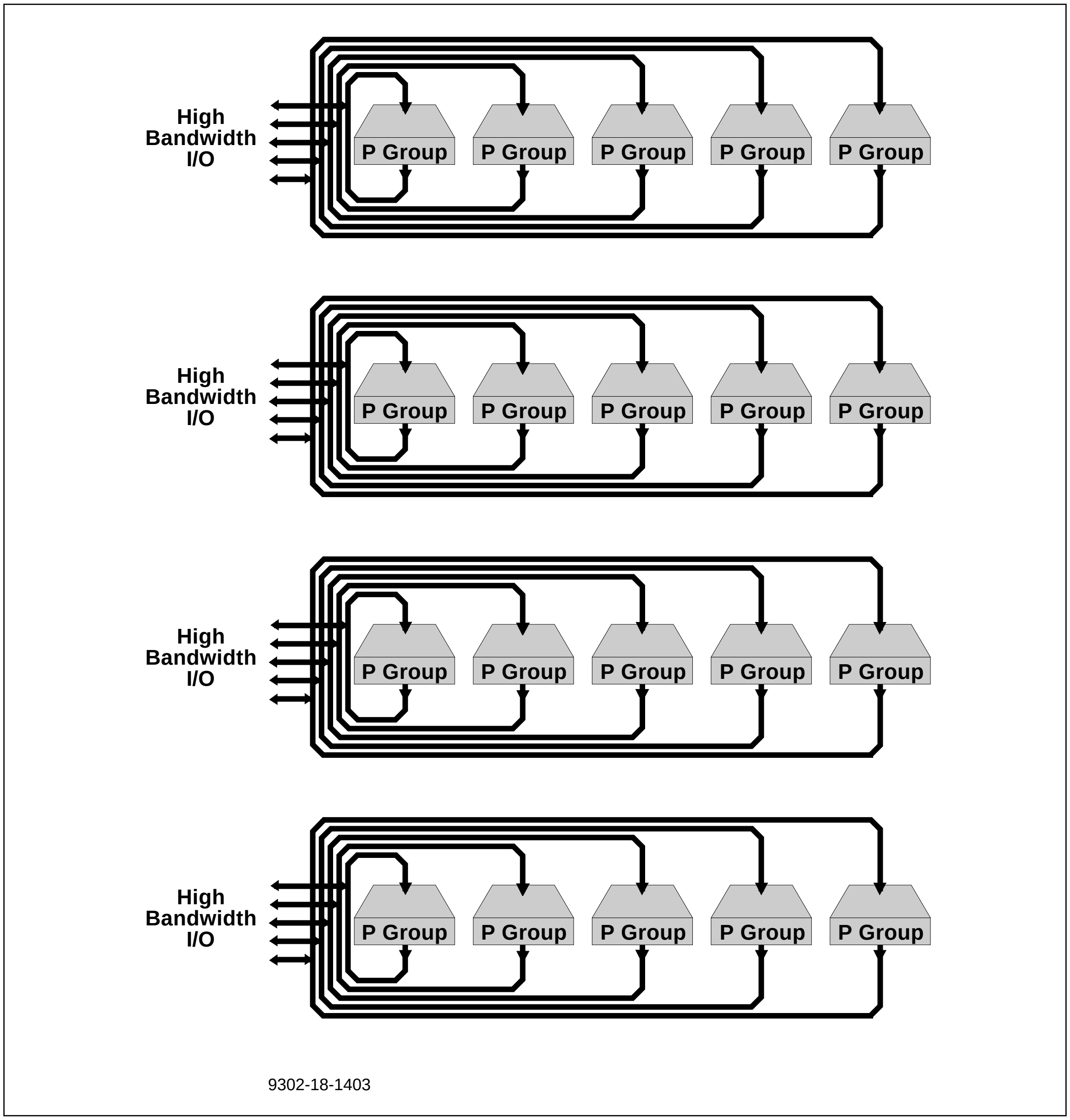


Figure 3 Type P Array with 20 Drive Groups

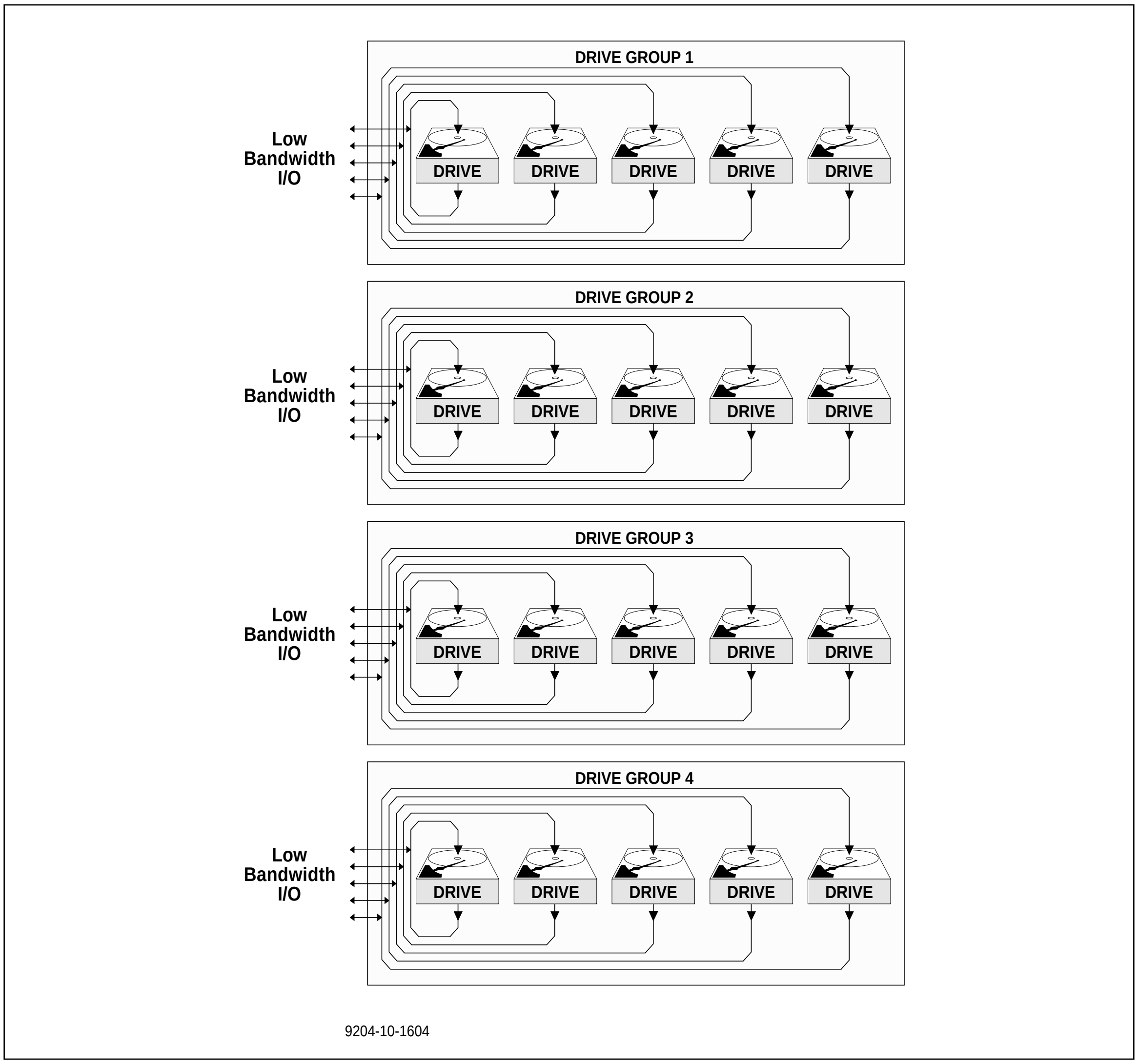


Figure 4 **Type S Array with 20 Drives**

We will refer to drives used in type P drive groups as P-drives and those used in type S drive groups as S-drives.

An important characteristic of the illustrations in Figure 3 and Figure 4 is that, at-a-glance, you can see the number of concurrent I/O data transfers possible and whether they are high or low bandwidth.

With knowledge of the drive's characteristics, we can determine upper bounds on performance of drive groups and arrays. For example, consider the specific type S drive group illustrated in Figure 2 which contains 5 drives. If each drive is capable, on the average, of 10 low bandwidth I/Os per second, the maximum number of low bandwidth I/Os per second for the entire drive group is 50. This is an upper bound because there is a very large amount of performance overhead associated with type S drive groups, and the actual performance of a type S drive group will be substantially less than the upper bound (See ECC Technologies' RAID & ECC Seminar material to understand the R/M/W performance overhead problem with type S arrays). If the same drives were used in the type P drive group illustrated in Figure 2, it would be capable of 10 high bandwidth I/Os per second.

Consider the upper bound on performance of the type P and type S arrays illustrated in Figure 3 and Figure 4. The upper bound on performance for the type P array in Figure 3 with 20 drive groups is 200 high bandwidth I/Os per second.The upper bound on performance for the type S array in Figure 4 with 20 drives is 200 low bandwidth I/Os per second. The actual performance limit of the type P array is very close to its upper bound, but the actual performance limit of the S array is significantly less than its upper bound.

The upper bound on drive group and drive array system performance as seen from the type of pictures shown in Figure 3 and Figure 4 does not account for the effect of disk array system controller design. Array system performance depends on the controller design in regard to the size of the cache, the algorithms used, and so forth. Figure 5 is a block diagram for a very high-performance disk array system which uses a high-performance multi-port memory and a high-performance crossbar (or crosspoint) switch. This type of controller design will provide the best possible performance.

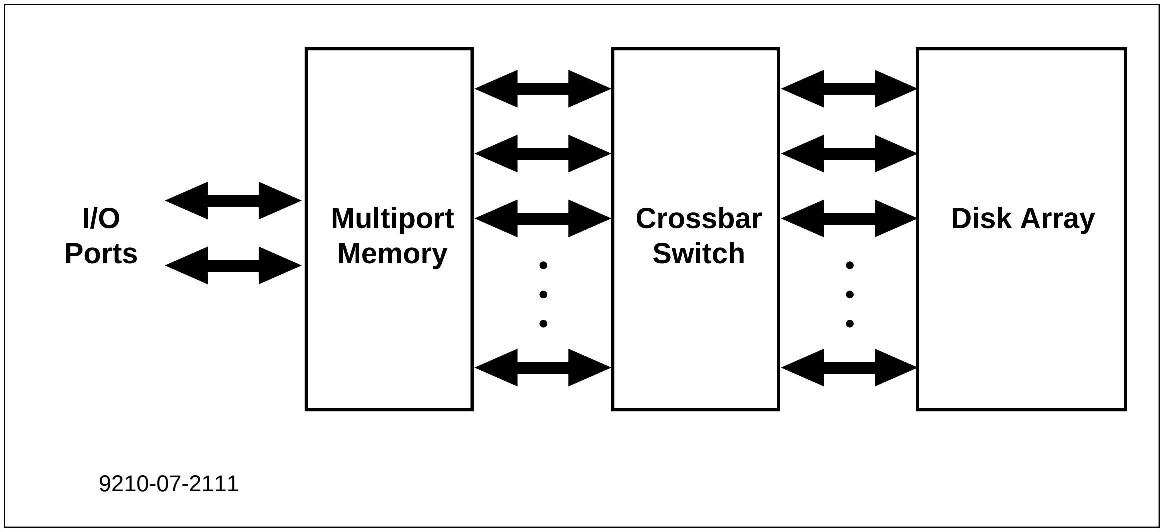
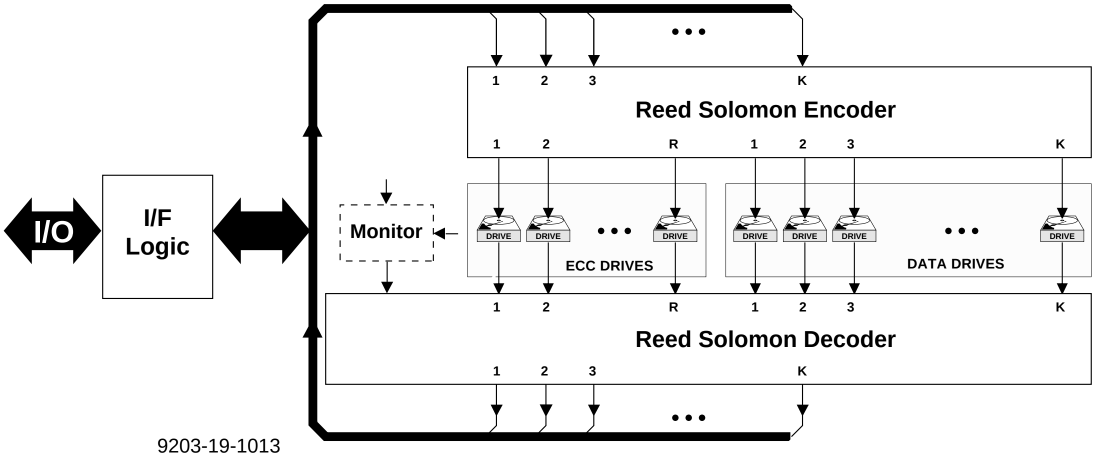


Figure 5 High Performance Disk Array System

2.1 P-Drives - Custom Drive Memory Elements

Conventional disk drives are serial, stand-alone devices. Conventional drives can be used in type P arrays, but using conventional drives with type P arrays creates a number of inefficiencies. These inefficiencies can be eliminated by designing a custom drive for type P arrays. We will refer to custom P-drives as “drive memory components” since these drives should not be designed for use as stand-alone devices but should be designed to be used as components or memory elements in type P drive groups.

Standard drive interface logic can be added to a type P drive group to create a parallel drive storage unit (PDSU). A typical Reed-Solomon (RS) PDSU is illustrated in Figure 6.



Monitor is optional

Figure 6 Reed Solomon (RS) Parallel Drive Storage Unit (PDSU)

ECC Technologies' primary focus is on P-drives. It is our belief that tiny-dumb drives will eventually be used as P-drives in place of conventional disk drives. We are firmly convinced that P Arrays will dominate the array market once tiny-dumb drives become widely available.

2.2 S-Drives - Conventional, Production Drives

S-drives can be conventional disk drives with conventional interfaces. Type S drive arrays in existence today use standard, production disk drives. There probably is nothing to be gained by designing a special type of drive for type S drive groups since standard drives have the features necessary for S-drives.

3. Reasons for Focusing on Tiny Drives

When we look at the overall trends in the information processing industry, we see the emergence of a new type of information processing architecture or topology as illustrated in Figure 7.

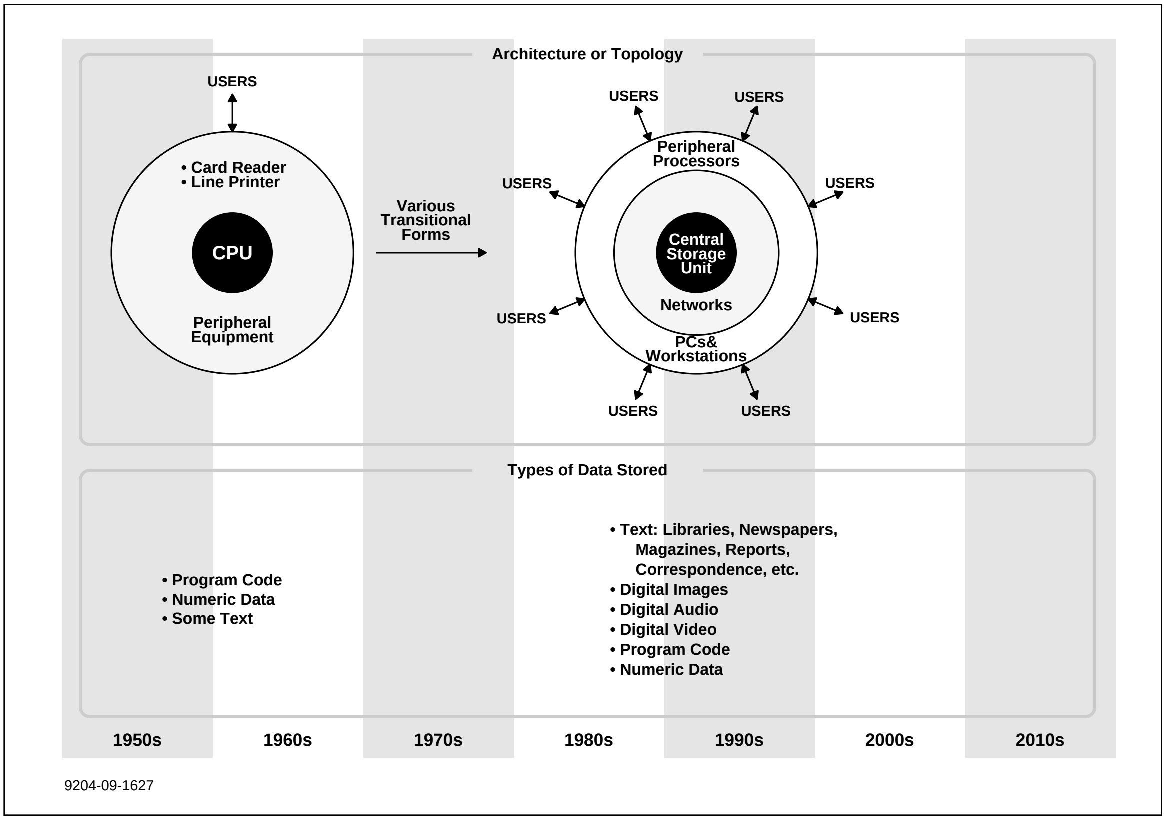


Figure 7 Evolution of Information Processing Systems

We see the emergence of multimedia PCs, interactive video, and massive-capacity databases accessible through high-performance networks. Many applications require a very large number of high bandwidth I/Os per second.

These ideas are not just abstract concepts. Many companies, such as Bell Atlantic, Southwestern Bell, American Telephone & Telegraph, GTE Corp., BellSouth, Ameritech, Tele-Communications Inc., Bell Communications Research Corp., Nynex, US West, Northern Telecom Ltd., and others are investing heavily in advanced cable and fiber networks so they will be capable of delivering the types of interactive information identified in Figure 7. Other companies, such as IBM, Time Warner and AT&T are investing heavily in service businesses that will deliver interactive information to end users with multimedia PCs. These companies understand the enormous profit potential for companies involved with these types of systems. It is the clear direction of the future.

The most logical way to achieve both high bandwidth and a large number of I/Os per second from drive arrays is to use type P arrays with a large number of type P drive groups. But, in order for this to be practical, each drive group must consume a relatively small space. This can only be accomplished if the drives are very small (or tiny). An extreme illustration of this would be the use of multiple IBM 3390s as P-drives. If we used 72 IBM 3390s in each type P drive group of a type P array, one drive group would take an entire room! On the other hand, if we used 5 mm thick 1.8” PCMCIA type drives, 72 drives would fit into the space of a small shoe box and not consume any more space than a conventional 5.25” drive. We choose 72 as an example because the standard SECDED code used with semiconductor memory contains 64 bits for data and 8 bits for redundancy. (“SECDED” stands for **S**ingle **E**rror **C**orrecting, **D**ouble **E**rror **D**etecting Hamming code.)

Simulation studies at several companies have shown rotational latency to be the key performance parameter affecting the performance of disk array systems. Tiny drives hold out the potential for smaller rotational latencies because they necessarily must be spun faster in order to maintain the same velocity of the head with respect to the media as conventional drives.

It is easier to spin tiny disks faster than large disks. The inertia of a disk is proportional to the fourth power of radius, and there is less surface area and less windage drag.

Figure 8 and Figure 9 compare the performance of type S arrays with N drives with the performance of type P arrays with N drive groups. Figure 8 shows the performance comparison when the I/O workload calls for the transfer of mostly long records and Figure 9 shows the comparison when the I/O workload contains mostly short records.

As drives get tinier and tinier, their capacity necessarily decreases. The 5 mm 1.8” drive mentioned above may contain only 50 MegaBytes. Forty of them ganged together would only contain 2 GigaBytes. Thus, it is a “fair” comparison to compare a type P drive group with 40 tiny-dumb drives as memory components to a single, conventional, larger drive such as a 2 GigaByte 3.5” drive.

It is easy to see from Figure 8 and Figure 9 that **type P drive arrays that contain the same number of drive groups as the number of drives in an equal capacity type S array will ALWAYS outperform the type S array by a wide margin**. And there will be an even more dramatic difference if tiny-dumb drives are made faster (smaller access times). **This fact may be the most important fact regarding drive arrays**.

This section has summarized some of ECC Technologies’ reasons for focusing on type P arrays and on fast, tiny-dumb drives. The next section will discuss ECC Technologies’ Reed-Solomon (RS) RAID technology. RS RAID technology is extremely important because two RS RAID schemes are capable of emulating every other existing RAID scheme in addition to providing a variable level of fault-tolerance.

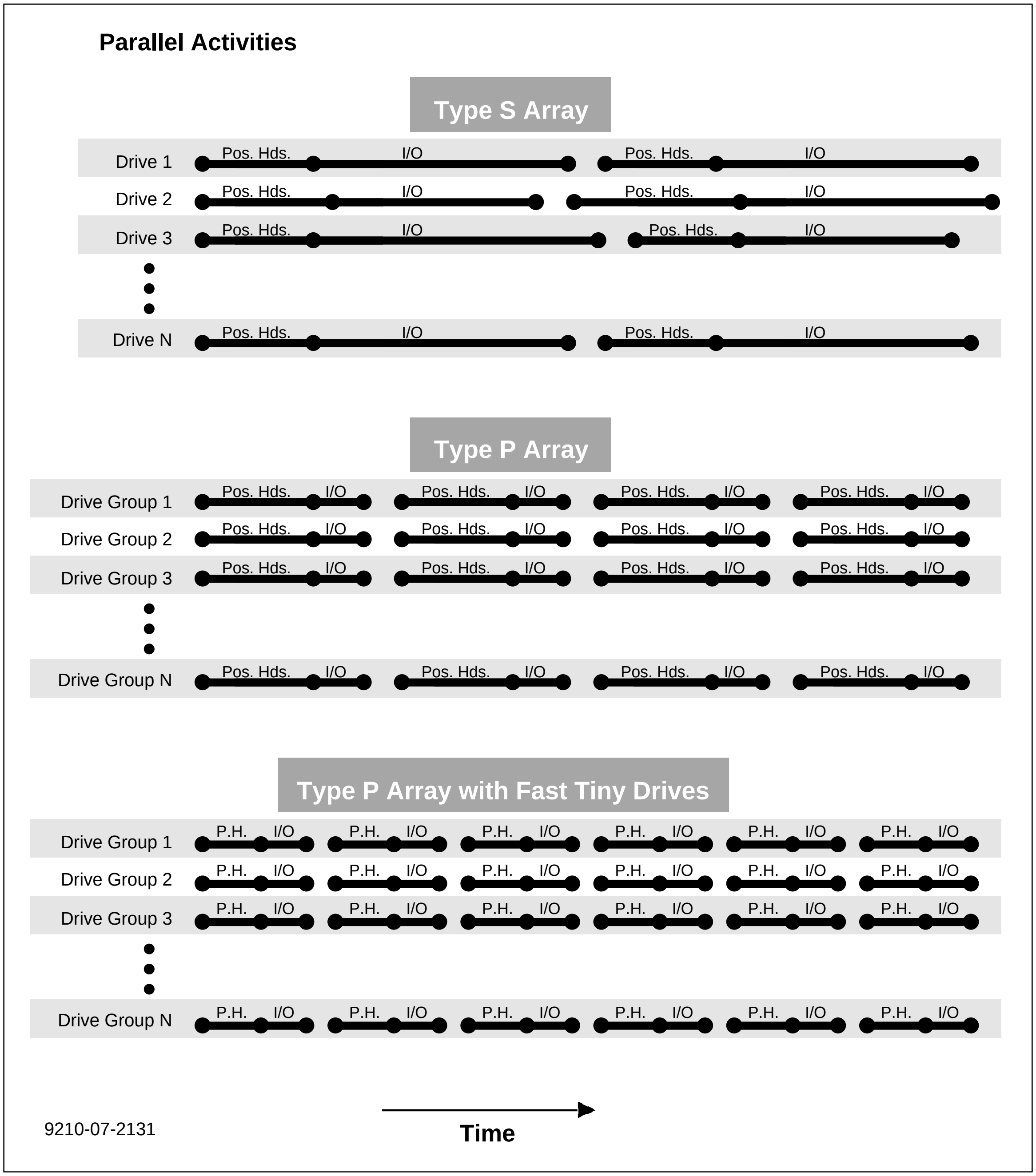


Figure 8 Performance of Arrays with Long Records

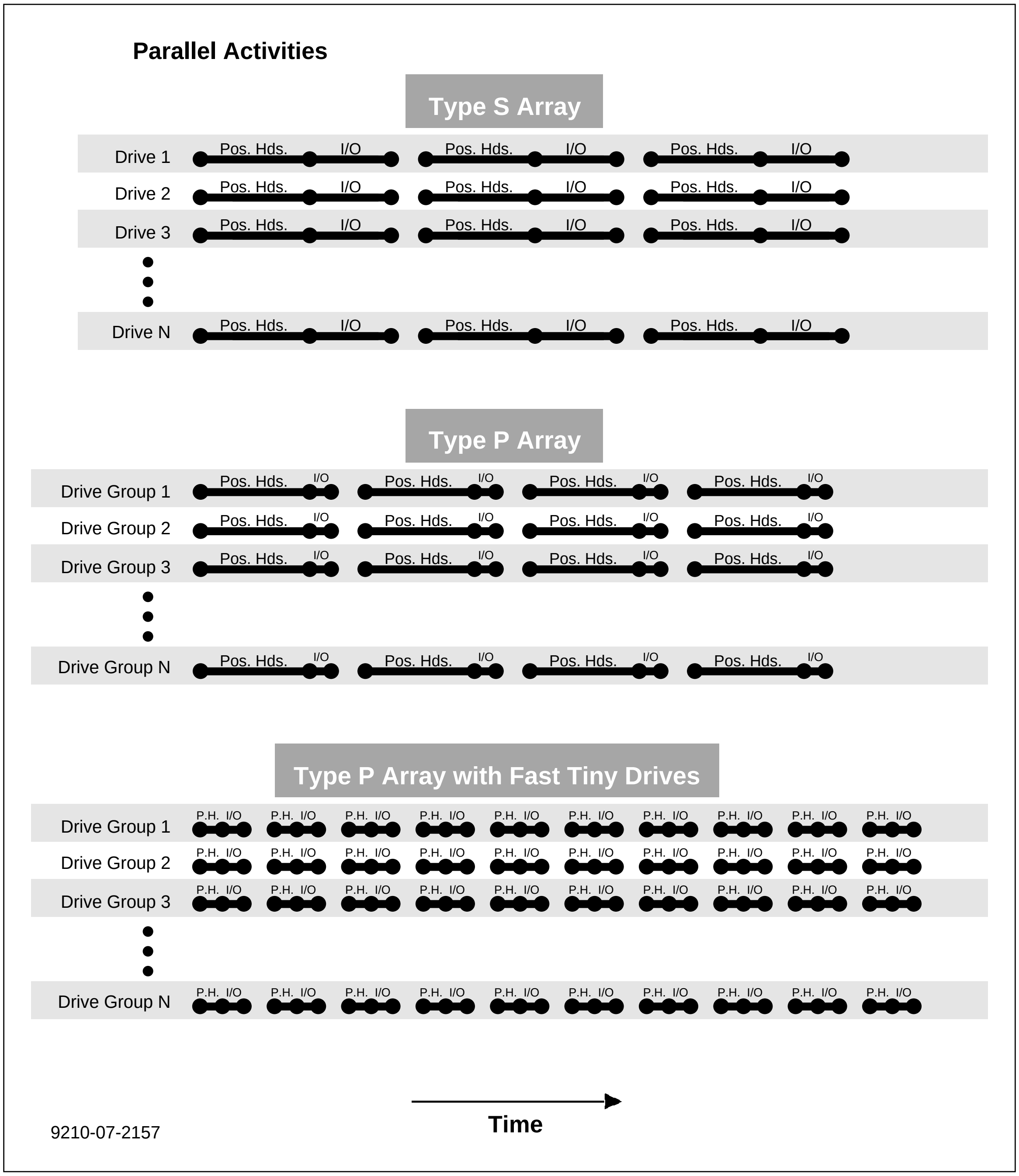


Figure 9 Performance of Arrays with Short Records

4. Reed-Solomon (RS) RAID

Drive arrays which contain redundant, parity-check information are referred to as **R**edundant **A**rrays of **I**nexpensive/Independent **D**rives or “**RAID**” for short. The acronym RAID was coined by researchers at the University of California in Berkeley in a 1987 research report which has come to be known as the 1987 Berkeley RAID Paper. In this paper, five RAID schemes were defined and identified as RAID 1, 2, 3, 4 and 5. These schemes were referred to as “levels”. ECC Technologies uses the word “schemes” instead of “levels” since “levels” erroneously implies a higher numbered level being more advanced than a lower numbered one.

Parallel, Reed-Solomon error correction technology can be applied to drive arrays in two fundamental ways. In one way, codeWORDS are striped across bytes from sectors in type P drive groups and, in the other way, codeWORDS are striped across bytes from sectors in type S drive groups.

When Reed-Solomon codeWords are striped across bytes from sectors in a type P drive group, ECC Technologies calls this type of RS RAID implementation “**RAID X**”. The primary characteristics and most severe problems of RAID X are listed in Figure 10.

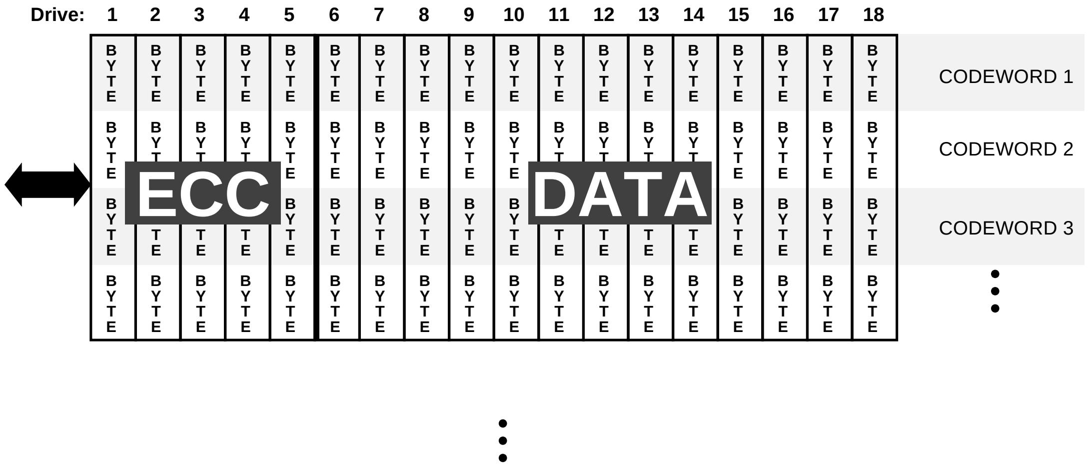
When Reed-Solomon codeWORDs are striped across bytes from sectors in type S drive groups, ECC Technologies has called this type of RS RAID implementation “**RAID Y**”.The primary characteristics and most severe problems of RAID Y are listed in Figure 11.

RAID X and RAID Y are the only RAID schemes drive array designers need to consider in designing any type of drive array. They are capable of emulating every other RAID scheme. What differentiates them from the other schemes is their ability to allow for a variable amount of redundancy per codeWORD which leads to a variable level of fault-tolerance.

This document focuses on RAID X because ECC Technologies strongly believes the R/M/W performance overhead inefficiency problem with RAID Y and all the other types of RAID schemes that use type S drive groups will eliminate them from serious consideration for next-generation disk array designs. The performance problems with type S arrays become extremely serious and troublesome if the drives in the array are unreliable or have high error rates. With RAID X, there is essentially no overhead performance penalty, all of the encoding and decoding is done on-the-fly in a single chip, and the performance of RAID X arrays is not affected by unreliable drives or by drives with high raw error rates as illustrated in Figure 12. (The reader is referred to the February 1992 special winter issue of ***Computer Technology Review*** for a more thorough comparison of RAID X to the other RAID schemes.)

RAID X

Data is written to and read from Sector Groups



PRIMARY CHARACTERISTICS

- An entire group of sectors is treated as a single record with K sectors for data and R sectors for redundancy (a Drive Group logically looks like a single drive)
- Uses a Reed Solomon code that corrects S soft error bytes and H hard error bytes in each codeword so long as $2S + H \leq R$
- Corrects error bytes caused by drive failures, defects, seek errors and random noise
- A single drive failure causes one hard error byte in each codeword
- All error correction done on-the-fly in hardware
- Allows a very high level of parallelism (High Bandwidth)
- Targets small-form-factor disks as memory elements
- ECC in the drives is not required and usually not desired

MOST SEVERE PROBLEMS

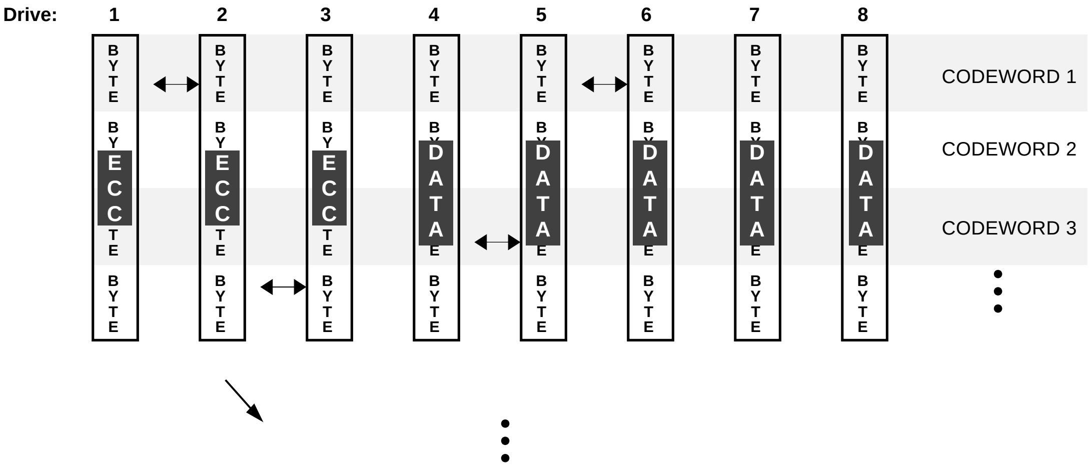
- In order to keep disk records relatively small, sector sizes must be small
- Small sector sizes reduce the disk formatting efficiency
- The number of I/Os/sec is limited by the number of Drive Groups in the I/O system

9108-05-1638

Figure 10 Characteristics and Problems of RAID X

RAID Y

Data is written to and read from Sector Groups



PRIMARY CHARACTERISTICS

- Sectors are read independently
- R sectors of redundancy and K sectors of data
- Sectors are written independently but there is a ***significant*** write overhead inefficiency
- Uses a Reed Solomon code that corrects S soft error bytes and H hard error bytes in each codeword so long as $2S + H \le R$
- A single drive failure causes one hard error byte in each codeword

MOST SEVERE PROBLEMS

- ***Significant*** write overhead inefficiency
- Further increased controller firmware complexity over RAID 5
- Severe degradation in performance when writing or reading a failed drive

9210-06-1530

Figure 11 Characteristics and Problems of RAID Y

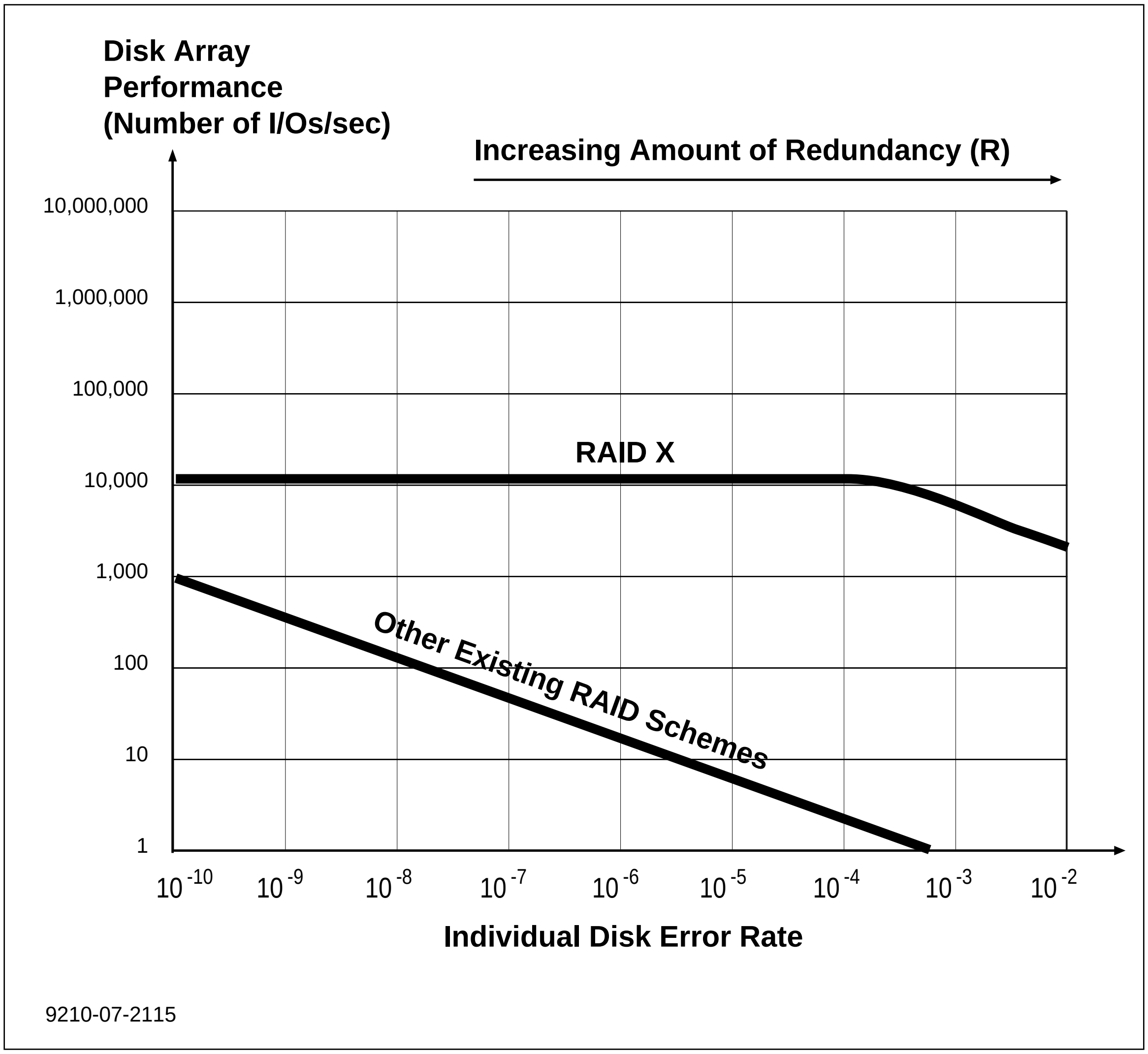


Figure 12 Disk Array Performance as a Function of Individual Disk Error Rate

The next section considers different types of recording objects and heads that are candidates for use in tiny drives. We will see that, for miniature recording devices, we must do some original thinking and reject some common myths.

5. Recording Objects & Heads for use in Tiny Drives

There are numerous different types of recording objects we have considered for use in tiny drives such as cylinders and nested cylinders, but, in order to save time and space, this document focuses on “disks” and “rods” since ECC Technologies believes they are the most likely candidates for use in tiny drives. Figure 13 shows a disk and a rod that have equivalent surface areas.

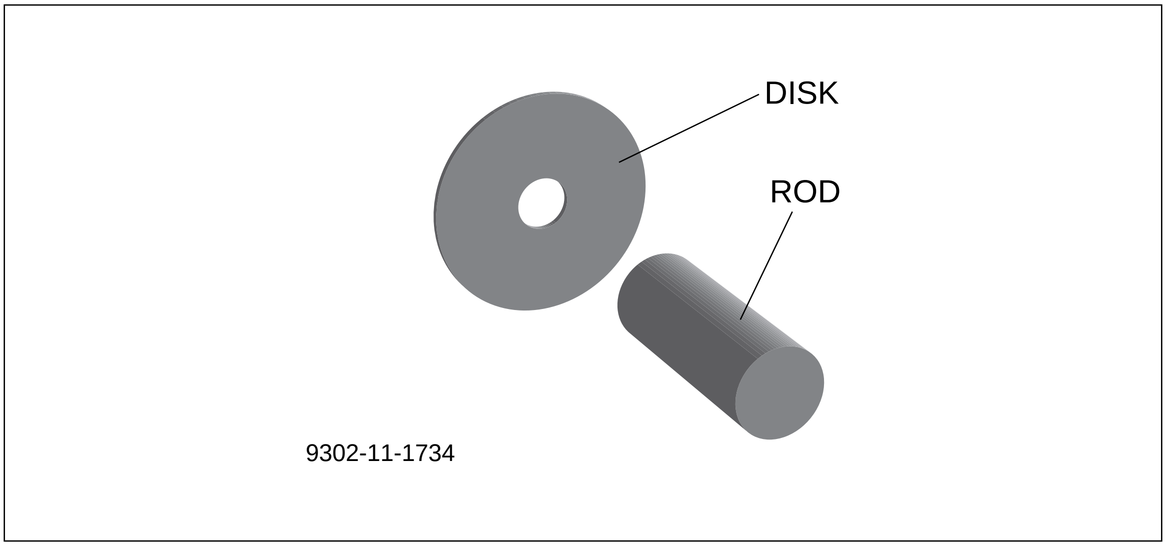


Figure 13 A Disk and Rod with the Same Recording Surface Area

In order to determine the “best” choice of a recording object for tiny drives, we have defined the “Efficiency” of recording objects. The efficiency of a recording object is the number of square units of recording surface area on that object divided by the number of cubic units of space it consumes. The spacial overhead incurred in an actual drive due to motor, actuator, heads and electronics is initially ignored in order to get a feeling as to which objects are most efficient when the geometries are scaled down into the “tiny” region.

The total surface area of a disk is $2\pi r^2 = 0.5\pi d^2$ where r is the radius of the disk and d is the diameter. If we assume the outer radius of the recording band is r and the inner radius of the recording band is $r/2$ (which is conventional), the recording surface area of a single disk is $1.5\pi r^2 = 0.375\pi d^2$. The slope of this curve is $0.75\pi d$. The slope is continuously increasing but is 0 when $d = 0$ as can be seen in Figure 14. This figure shows the recording surface area of a disk as a function of its diameter for small diameters. We may say the recording surface area of disks increases very slowly with diameter if we start looking at disks with diameters near 0 and then consider larger diameter disks or we may say the recording surface area of disks decreases very rapidly with diameter if we start looking at large diameter disks and then consider using smaller diameter disks.

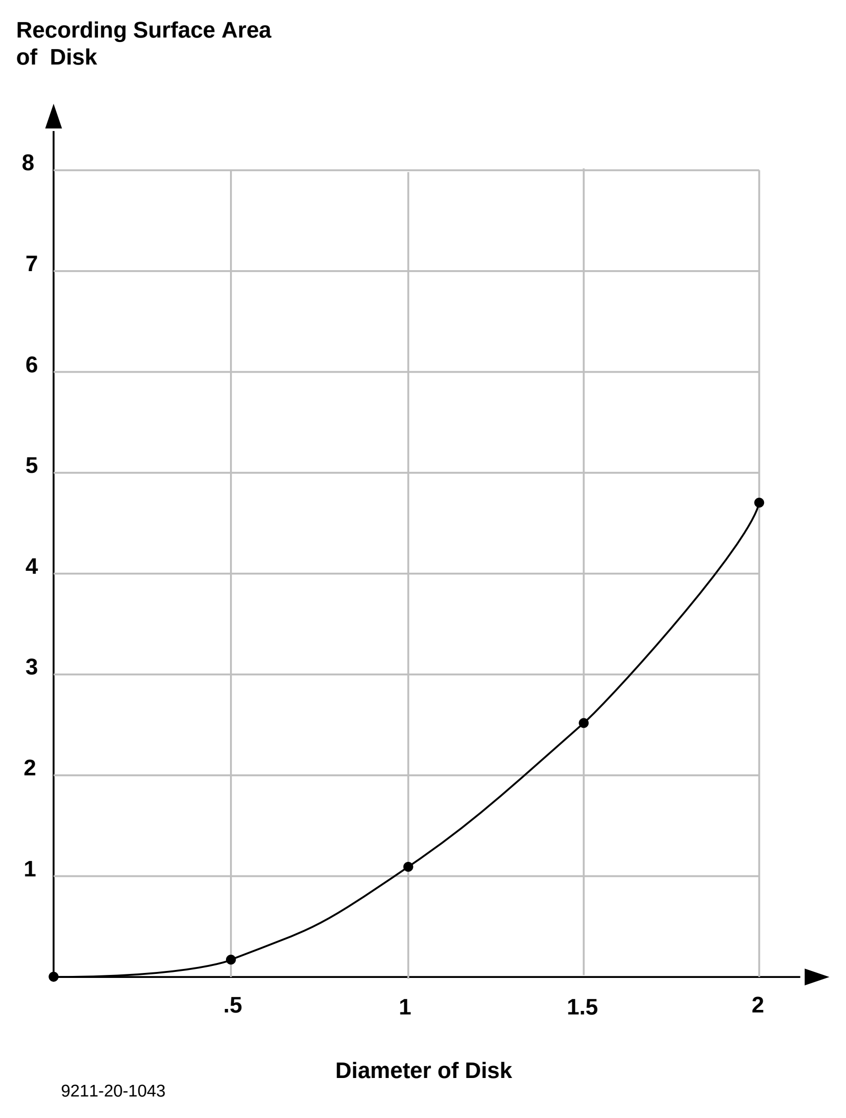


Figure 14 **Recording Surface Area of Disks as a Function of Disk Diameter**

The surface area of a rod is πld where l is the length of the rod and d is the diameter. This is a straight line at constant slope of πl. The surface area of rods of length 0.25, 0.5 and 1 are shown in Figure 15.

Figure 14 and Figure 15 are superimposed in Figure 16 so the difference in recording surface areas between disks and a rods can be clearly seen. Rods contain more surface area than disks at small diameters.

Next, we compare the efficiency of disks to the efficiency of rods.

If we assume the inner radius of a disk recording band is half the outer radius, the efficiency of a disk is $1.5/t$ where t is the thickness of the disk. The disk efficiency does not

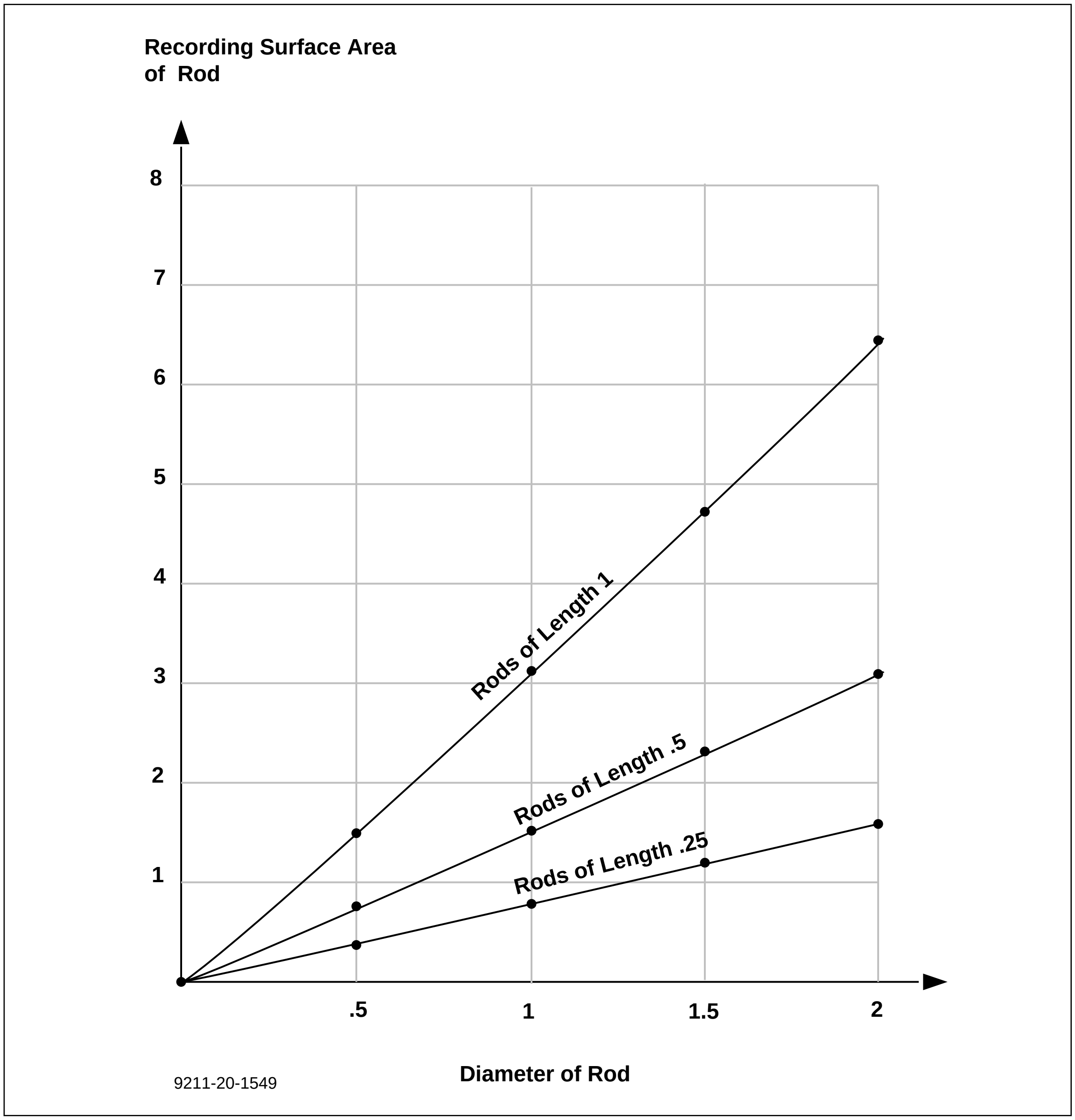


Figure 15 Recording Surface Area of Rods as a Function of Rod Diameter

depend on the radius - only the thickness. As the thickness of the disk goes to 0, the efficiency of the disk goes to infinity. The efficiency of a disk as a function of thickness is shown in Figure 17.

The efficiency of a rod is $4/d$ where d is the diameter of the rod. In the case of a rod, the efficiency only depends on the diameter of the rod and not on its length. The efficiency of the rod as a function of its diameter is shown in Figure 18.

As an example, a disk of thickness 0.25 has an efficiency of 6 and a rod of diameter 0.5 has an efficiency of 8. In this case, the rod is more efficient than the disk. Figure 19 compares the efficiency of rods with the efficiency of disks.

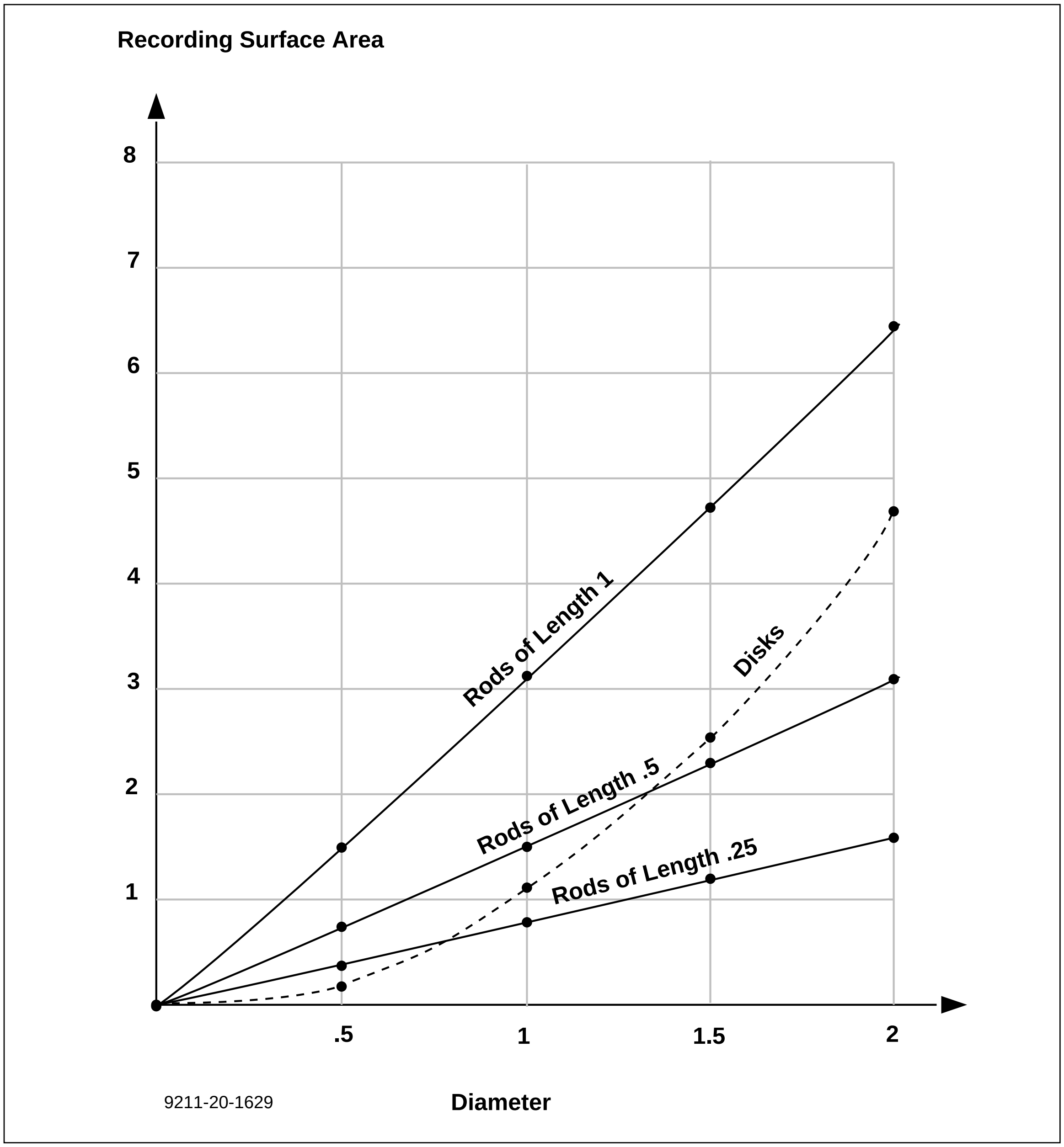


Figure 16 Surface Area of Rods Compared to Surface Area of Disks

Cylinders and nested cylinders are found to be the most efficient, but they are probably impractical to implement because they would require multiple types of sliders. It is concluded that “rods” are probably the “best” practical choice for recording objects at small diameters because they become more efficient as their diameter is decreased and only one new type of slider must be designed.

In the next section we briefly summarize the situation in regard to using conventional disk drive technology for tiny drives.

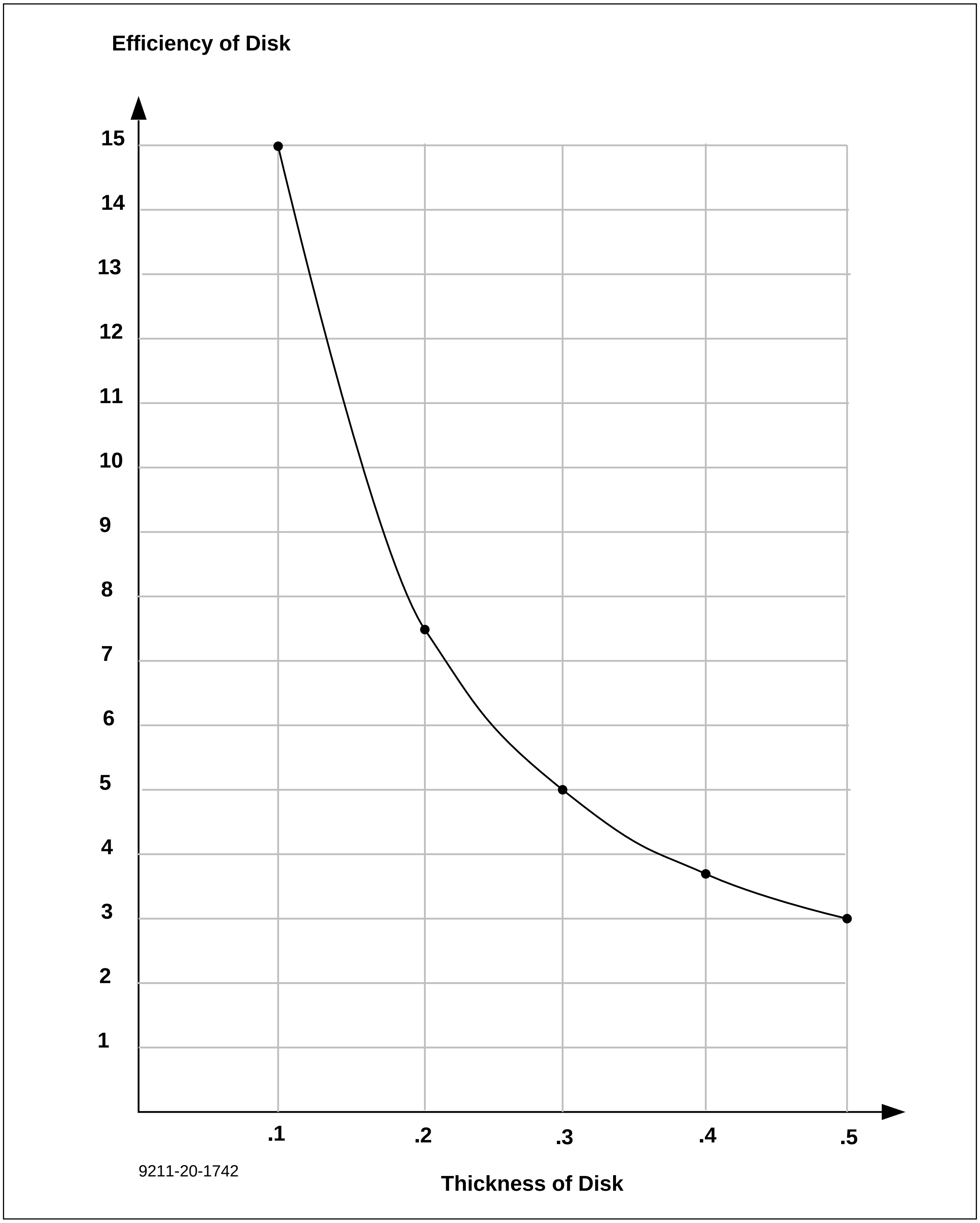


Figure 17 Efficiency of a Disk as a Function of Disk Thickness

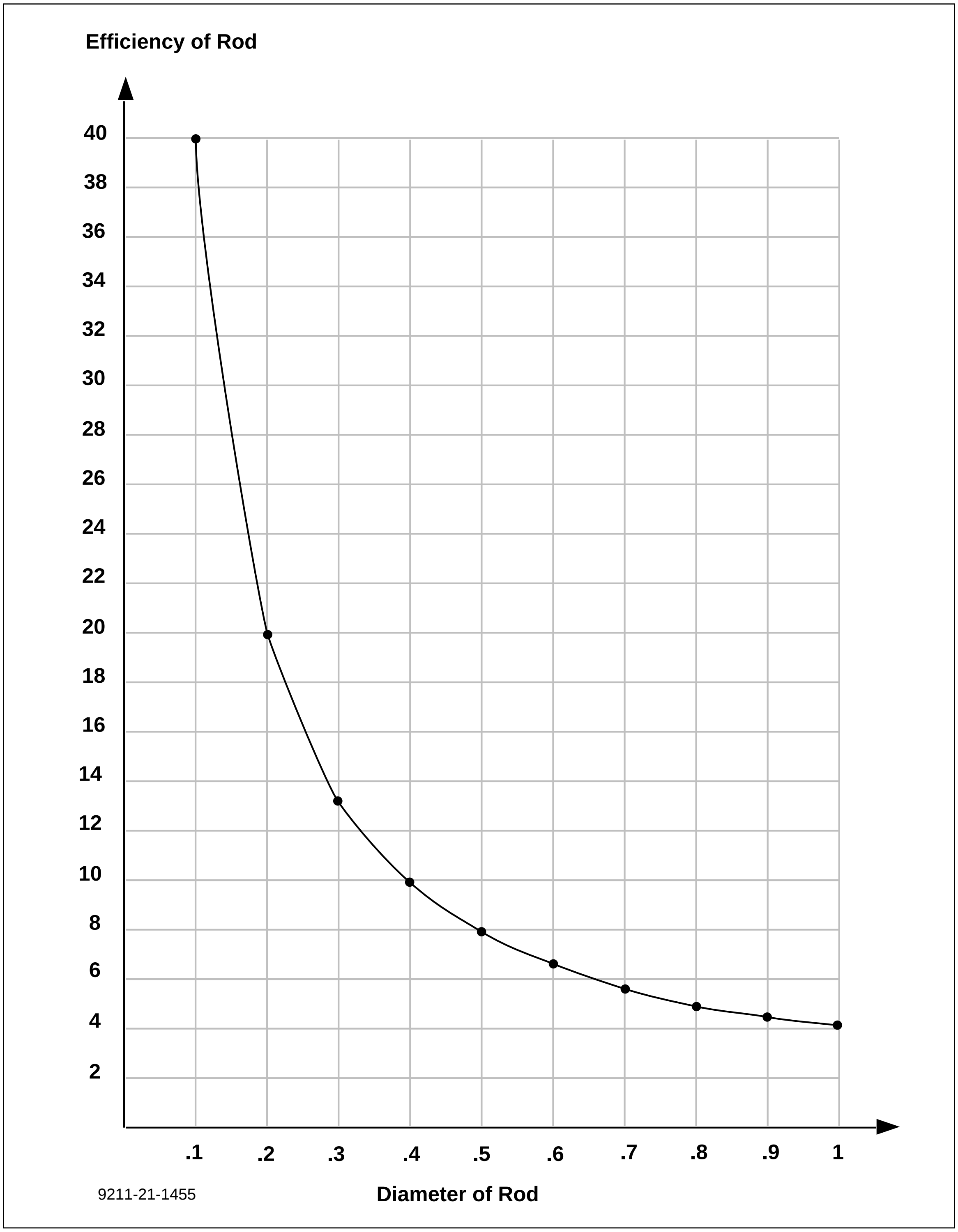


Figure 18 Efficiency of a Rod as a Function of Rod Diameter

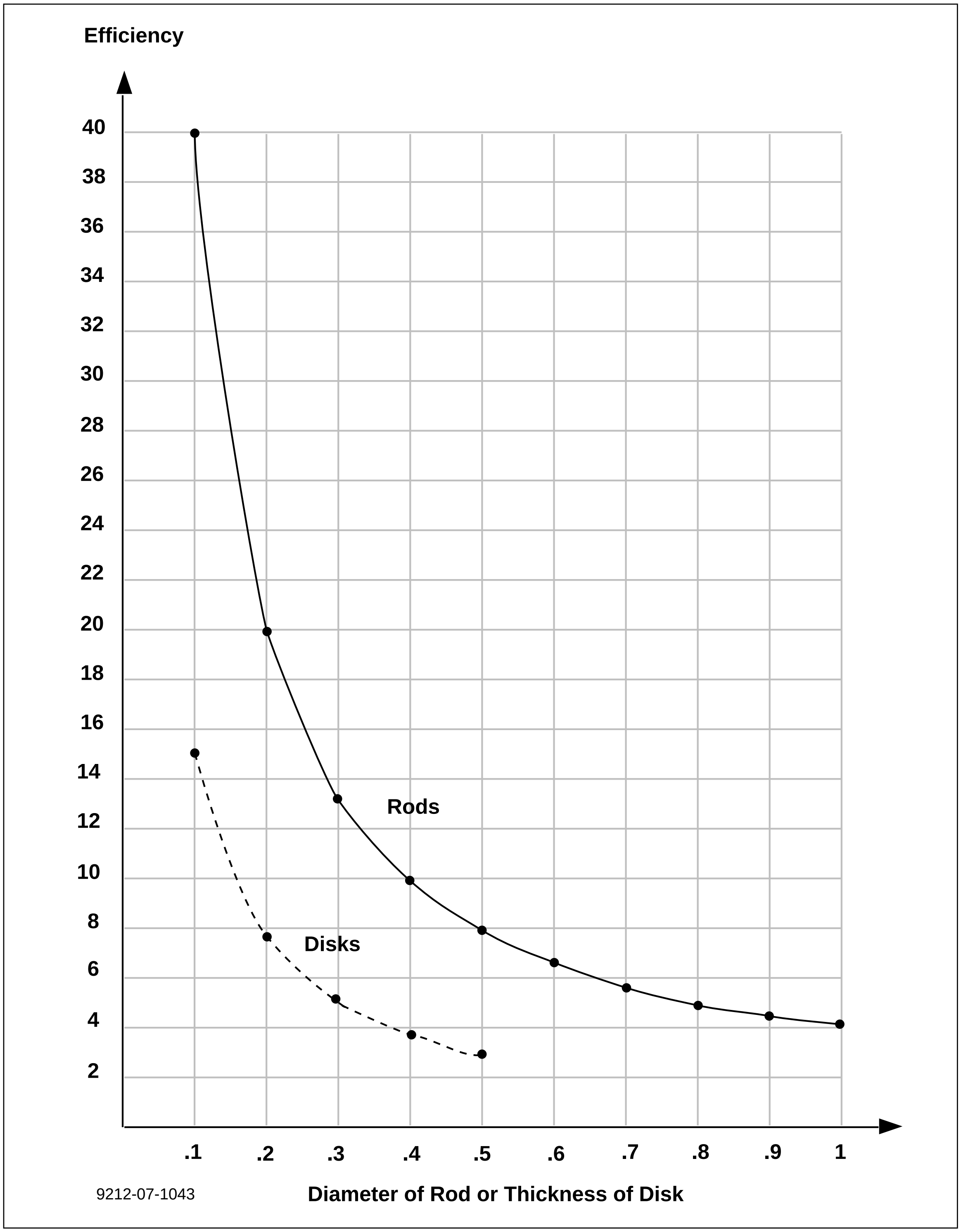


Figure 19 Efficiency of Rods Compared to Efficiency of Disks

6. Disk Drive Technology for Tiny Drives

In this section we take a look at conventional disk drive technology and see its limitations as disk diameters are scaled down to the inch or sub-inch range.

Disk drives were invented by IBM in San Jose, California around 1957. Before that time, large diameter drums were used as mass storage devices. Large diameter drums gradually disappeared from the scene since it was observed that if a large diameter drum is sliced like a loaf of bread, the resulting set of “disks” have a much larger surface area than the original drum. In other words, the resultant set of disks is more efficient in terms of the number of square units of recording surface area per cubic unit of volume consumed. The belief that all disks are more efficient than all drums has remained essentially unchallenged up to the present day.

The history of disk drives shows several trends. The most popular and highest volume drives have used smaller and smaller diameter disks as time went on. Disk diameters have gone from 24 to 14 to 12 to 8 to 5.25 to 3.5 to 2.5 and now to 1.8 and 1.3 inches. The average number of disks per drive has decreased, and the capacity per drive has steadily decreased as a result of going to smaller diameter disks with fewer disks per spindle. The number of MegaBytes per cubic inch has steadily increased with smaller diameter disks while the amount of power consumed per MegaByte has steadily decreased. These facts strongly reinforce ECC Technologies’ belief that future drive arrays will use arrays of tiny drives and that these tiny drives will be viewed as “drive memory components.”

As discussed in the previous section, very small diameter disks present a problem since the surface area is increasing so slowly with diameter or decreasing so rapidly with diameter - depending on how you look at it. To solve this problem we turn from disk drive technology to rod drive technology for tiny drives.

7. Rod Drive Technology for Tiny Drives

This section describes a new type of magnetic recording drive technology referred to as rod drive technology since the rotating recording object is a “rod” instead of a “disk” as is used in conventional magnetic disk drives. A rod drive is illustrated in Figure 20.

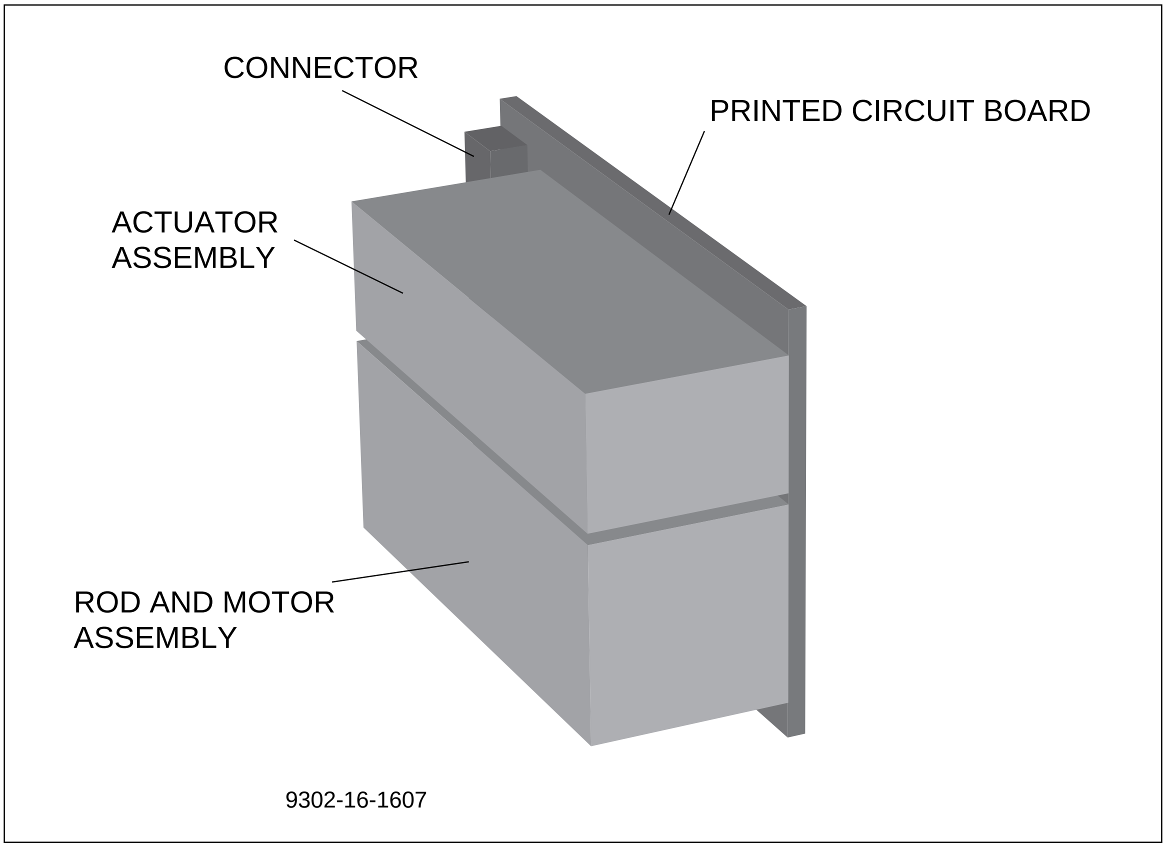


Figure 20 Rod Drive Illustration

At small diameters, rod drives have many advantages over conventional disk drives as summarized below.

Rod drives have a uniform recording density and a constant head flying height which simplifies the design of the recording channel electronics and enables the design of an adaptive-type of channel that compensates for pattern-dependent “peakshift”. That is, the channel characteristics are measured and postcompensation circuits are designed based on the channel’s characteristics. With rods, we only need to do this for one track since all tracks have the same characteristics. This is a very significant advantage over disk drives since, in disks, each track on a surface has slightly different density and flying height characteristics from all other tracks. Adaptive equalizers for disk channels must be programmed differently for each track (or zone) on a surface. For a rod, the adaptive equalizer needs to remember only one set of very precise characteristics.

The desired end-result effect of designing a sophisticated adaptive equalizer channel is the ability to trade off capacity (or density) for error rate. In other words, a very low raw error rate drive with a capacity factor of, say, 1, may be designed or a higher raw error rate drive with a capacity factor of, say 1.5 may be designed. We also may deliberately design a very high raw error rate drive with a higher capacity factor of, say 3. This means the high raw error rate drive is capable of holding 3 times as much data as the low raw error rate drive.

High raw error rate drives are perfectly acceptable in parallel drive storage units (PDSUs) or type P drive groups in disk arrays that use parallel Reed-Solomon error correction technology. Reed Solomon codes can handle drive raw error rates as high as 10^{-2}. (See ECC Technologies' RAID & ECC Seminar material for more information about error rates before and after decoding.)

Drive raw error rate as a continuous function of density is illustrated in Figure 21. The exact shape of this curve is not known, but it is our firm belief that it is possible to design channels to trade off error rate for density. Each point on the curve in Figure 21 will require the adaptive equalizer to be programmed differently.

It should be strongly stated that <u>designing a higher raw error rate drive is not the same thing as cranking up the frequency on conventional drives</u>. In conventional drives, raw error rate is a discontinuous function of density because the positions of detected flux transitions are often not normally distributed and are pattern-dependent. In order to properly design a high raw error rate drive, the channel electronics must be designed with extensive knowledge of the recording characteristics of the channel so that an adaptive equalizer may be designed. ECC Technologies proposes a channel equalizer referred to as a "context-dependent postcompensation system." (More details of the context-dependent postcompensation system are discussed in Section 8.2 on page 35.)

As disks get thinner and thinner, their diameter necessarily must decrease in order to maintain the necessary rigidity. Rods remain rigid even at small diameters and become more efficient as their diameter is decreased. Rods appear to be the ideal choice for miniature recording devices since their efficiency increases as their size is scaled down.

A rod has only one recording surface while a disk has two. Only one head is required for a rod, but two are required for a disk. This simplifies the design of rod drives, reduces the mass of the rod drive's R/W head-arm assembly compared to a disk drive and may cost less to manufacture than equivalent capacity disk drives. The head positioning systems in rod drives may take less power than head positioning systems in disk drives since they need to move less mass.

Rod drives should be much faster than disk drives in system applications because they probably can be spun faster than disks. Higher RPM leads to shorter rotational latencies. As mentioned earlier, simulation studies have shown rotational latency to be the most important performance parameter affecting system performance in disk I/O systems. High RPM rod drives used in future type P drive arrays will provide higher performance in terms of the number of high bandwidth I/Os per second possible.

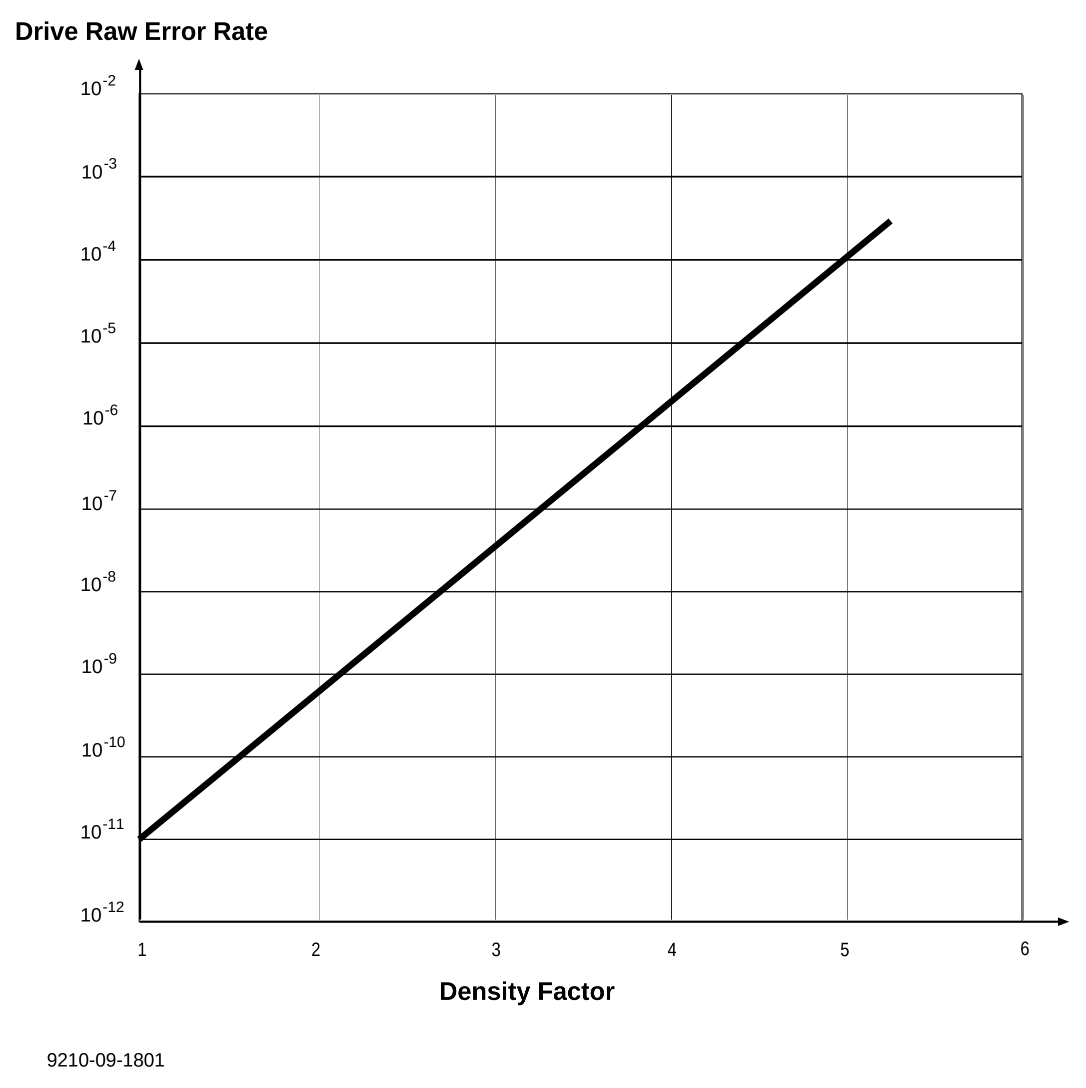


Figure 21 Drive Raw Error Rate as a Function of Density

It is possible with rod drives to design a ring slider with multiple heads per track to reduce the rotational latency of the drive even further than is possible by only increasing the RPM. With multiple heads per track in a ring, the head closest to the data can be electronically selected for reading or writing. The rod geometry allows a ring slider to be moved with a single actuator which is impossible with a disk drive. With a disk drive, multiple heads per track require multiple actuators. A ring slider is illustrated in Figure 22.

It is also possible with rod drives to design a stationary head-per-track bar slider as illustrated in Figure 23. With this arrangement, no actuator is required. This type of design is similar to the way some laser printers write onto photosensitive drums. We do not know if this is a “practical” concept or not. It is believed that the cost of the head-per-track bar

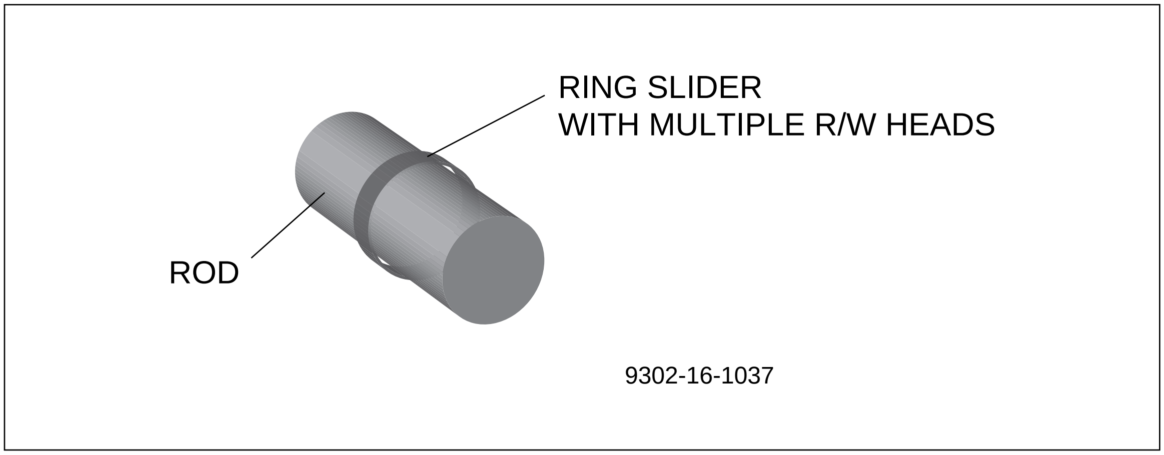


Figure 22 Illustration of a Ring Slider on a Rod

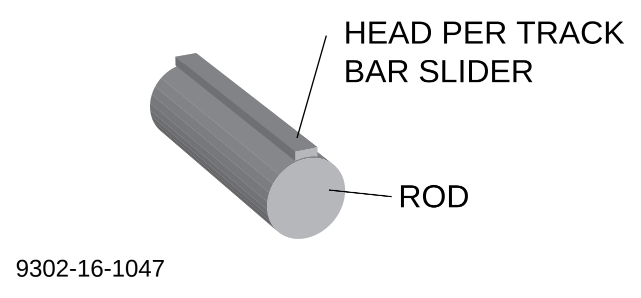


Figure 23 Rod with a Stationary Head-Per-Track Bar Slider

could greatly increase the cost of the device and may render the design impractical at this time.

Combining the ideas illustrated in Figure 22 and Figure 23 leads to a rod inside of a cylinder with multiple heads per track. This is the way “drums” were designed in the late 40s! With this structure, either the cylinder or the rod can rotate while the other object remains stationary.

Rods are more rigid than disks. This extra rigidity makes it possible to design rod drives with higher mechanical precision. Higher rigidity also translates into more rugged devices.

7.1 Variations on the Rod Drive Concept

This document focuses on rotating rod devices, but many of the ideas also apply to other types of unconventional recording objects such as cylinders and nested cylinders. In those cases the heads may rotate instead of the media. Spinning rods are probably the most cost-effective and practical since only one type of head has to be used. Cylinders and nested cylinders require two or more types of heads.

7.2 Rod PDSUs

Parallel Reed-Solomon error correction technology can be applied to a set of tiny rod drives to form parallel drive storage units (PDSUs). Rod PDSUs have a number of advantages over conventional large-capacity disk drives. High RPM rod drives lead to shorter access times, the parallel configuration leads to higher transfer rates and higher error rate rod drives will allow more MegaBytes to be stored per square inch of recording surface area. Since rods are highly efficient at small diameters, small diameter parallel rod drive storage units will dramatically increase the number of MegaBytes stored per cubic inch of volume over conventional disk drives. Rod PDSUs are also fault-tolerant so that one or more rod drives can fail with no loss in performance or data.

Rod PDSUs should be able to compete favorably with high capacity disk drives. Consider a 2 GigaByte conventional drive compared to a 2 GigaByte Rod PDSU. The Rod PDSU may initially cost more, but has the potential to cost less if it becomes commonplace and the economies of scale kick-in. Even at a higher cost, the benefit to the user far outweighs the cost. Higher transfer rate, faster access time, fault-tolerance and built-in backup increase user productivity and sense of security in knowing the data is correct and complete. In some industries data security and/or fast data access is crucial to a successful business. As the future demands more and more information delivered “right now”, the Rod PDSU will be a perfect fit.

Rod PDSUs can be designed with fixed media or removable media. With fixed media rod PDSUs, each rod is permanently attached to a motor. With removable media (Rod Cartridge) PDSUs, a set of rods (similar to a set of quick-connect drill bits) would probably be contained in a plastic enclosure with compartments for each rod. Insertion of the cartridge causes the rods to be extracted and quick-connected to a set of tiny motors. Heads are then loaded onto the rods for writing and reading.

An alternative Rod PDSU Cartridge would use a sliding metal plate with multiple windows - one window for each rod. This design is similar to the design of 3.5” floppy diskettes. The diskettes contain only one window per plate; the rod cartridge would have multiple windows per plate.

8. Tiny Drive Electronics

ECC Technologies is promoting what it calls “tiny-dumb drive technology” which can be used to design very efficient P-drives for RAID X drive groups. This document focuses on the design of circuits for tiny-dumb P-drives. The reader should keep in mind that most of these ideas also apply to conventional disk drives.

RAID X is most efficient if the number of P-drives in a drive group is relatively large. If the number of P-drives in a group is relatively large, the amount of redundancy necessary for error correction is small. Coding theory tells us that long codeWORDs are best, but, a large number of P-drives per group causes a formatting inefficiency problem if record sizes are relatively small. In order to solve this problem, ECC Technologies has invented a fast Phase-Lock Loop (PLL) synchronization system that we will refer to as “Fast Sync”.

In addition to advocating the Fast Sync system, we are advocating the elimination of headers (or address fields) and ECC fields in drive formats and the elimination of ECC circuits in the drive.

ECC Technologies has studied the problem of making the raw error rate a smooth, continuous function of density or capacity as illustrated in Figure 21. A superstition has existed in the disk industry for several decades that if density is increased beyond a certain point, the error rate will “go off a cliff”. More accurately, we should say the error rate would “go up a cliff”. We believe this superstition is similar to the superstition that said the world is flat and that if you went too far out into the ocean on a boat, you would eventually “go over the edge”. What has fueled this superstition is that many attempts have been made to increase the density of disk drives at the expense of raw error rate by increasing the frequency of the clock only. These attempts will always fail because the positions of detected transitions within a detection interval are usually not normally distributed when using conventional disk channel electronics and deriving timing information from the signal is a significant problem when using conventional PLLs.

It is imperative that, in order to achieve the objective of making raw error rate from a saturating magnetic recording channel a smooth function of density, we must come up with a system to “randomize the noise” by making the distribution of detected flux transition positions more like a very steep normal distribution where the most likely position is in the center of the window. Such a system attempts to remove pattern-dependent (or context-dependent) position shift. Circuits that accomplish this goal are similar to image enhancement circuits or algorithms in that they enhance good signal areas at the expense of possibly degrading bad signal areas. As a result of this, we know a priori that the resulting distribution of positions will probably not be perfectly normally distributed and arguments for or against these types of schemes based on an assumption that the positions are normally distributed are probably not valid arguments.

The following analogy should clarify what our objective is and what intermediate results we can expect. Imagine that you have a multi-page document that you take to a reducing copy machine. You copy 4 pages reducing them by 50%. You then cut and paste them onto a single page. You continue this process (quadrupling the storage capacity of a single page in each step) until it becomes impossible to read the text - even with a magnifying glass. The original document can be read with no errors, but, as we start reducing, the text becomes increasingly more difficult to read. Defects in the paper or in the copying process will eventually cause some letters to be obliterated, but we are often able to determine what words are by their context. When characters are reduced to the size of a toner particle, they become unrecognizable. This is a clear illustration of what we are attempting to do with magnetic recording devices - keep increasing the capacity at the expense of error rate until the recorded information becomes unreadable.

ECC Technologies has designed an adaptive equalizer for making raw error rate a smooth, monotonically increasing function of density. The equalizer requires the use of what we refer to as a “context-dependent postcompensation system”. With this system, the position of a target flux transition is adjusted based on its context. The system is based on the assumption that the flux position shift of a targeted transition is almost totally dependent on the position of its nearby neighboring transitions - that is, on its context.

The nature of this dependency relation can be determined by precise measurement. A channel characterization testing system is proposed which measures this time shift for each possible context. The results of this testing are used to determine how to program the programmable context-dependent postcompensation system. A first version of this type of tester has previously been designed and developed by ECC Technologies’ personnel.

8.1 Independent Clock Track & Fast Sync

ECC Technologies’ tiny-dumb rod drive contains a movable R/W data head and a stationary Read Only (RO) clock head as illustrated in Figure 24.

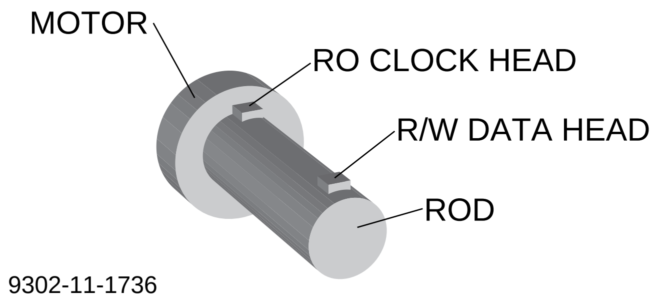


Figure 24 Rod & Motor Assembly Showing R/W Head and RO Clock Head

With an independent clock track, the frequency of the rod is always known and available to the drive’s electronics. To synchronize the clock signal to the signal from the R/W data head, we must make adjustments to the phase of the clock signal but not to its frequency.

The clock signal provides feedback to the motor control electronics simplifying the design of the motor by removing the necessity of having rotation position sensing circuitry designed into the motor. The Rod and Motor are one integrated assembly.

The clock signal provides continuous rod position information for use in the R/W head positioning electronics and for use in determining which sector the R/W head is over.

When writing, the clock signal, adjusted for phase, is the write clock. When reading, the clock signal, adjusted for phase, is the read clock.

The clock signal greatly reduces the lengths of the PLL synchronization fields and greatly simplifies the design of the read and write PLLs. Conventional PLL fields with from 12 to 15 bytes can be reduced to a few bits in length. With a clock signal, the drive's PLLs do not require variable frequency oscillators (VFOs). VFOs are replaced by high-resolution programmable delay lines which are very easily implemented in high-density digital ICs. The phase of the clock signal is adjusted by instantaneously reprogramming the programmable delay line.

To acquire lock, the PLL determines the phase error between the clock signal and the Read signal reading the PLL synchronization field. A number of phase error samples (preferable the number is a power of 2) is obtained, and the average of the samples is taken to be the true phase error. An instantaneous phase change is accomplished by placing an appropriate setting into the high-resolution programmable delay line. Thus, within a few bit times, we have acquired phase lock.

The frequency of the write or read clock signal is always the same as the frequency of the clock signal. The only difference is a possible phase difference.

Another advantage of having a clock track is that we do not have to use run-length-limiting (RLL) codes if we do not want to or we may choose to use a RLL code with a minium run-length constraint but no maximum run-length constraint. We may still desire to use (d,k) RLL codes to increase density, but we do not have to be worried about the k constraint. (the d constraint is the minimum run of non-transition symbols between transition symbols and the k constraint is the maximum run of non-transition symbols between transition symbols.) Drives with a RLL k constraint of ∞ will have a greater capacity than drives with a RLL k constraint of 7 such as the (2,7) or the (1,7) RLL codes commonly used in today's production disk drives. Bounds on the information density achievable with RLL codes are presented in an article in ***Information and Control*** entitled **Block Codes for a Class of Constrained Noiseless Channels**, Vol 17, pp 436-461 (1970).

An independent clock track opens the door to other methods for increasing the density of magnetic recording channels. One system under consideration by ECC Technologies is a non-saturating system that encodes data by doing an inverse Fast Fourier Transform and decodes data by doing a Fast Fourier Transform on the playback signal. Amplitudes and phases for a discrete set of frequencies within the passband of the channel are selected by the data to be encoded. The encoder performs an inverse FFT and creates a noise-like signal which is recorded. On reading, the decoder performs an FFT on the playback signal and recovers the original set of amplitudes and phases which determines the data. This system will definitely work if the disk channel is linear. Tests for linearity should be done on future recording channels to determine if this method is valid or not. If the channel proves linear or can be made to be linear, this method may dramatically increase the capacity of magnetic recording channels. This method has the potential to make full use of the recording capability of the channel, and raw error rate can be traded off for density by increasing the number of discrete amplitudes, phases and/or frequencies used with the scheme.

8.2 Context-Dependent Postcompensation System

The effect of postcompensation to the playback signal is similar to the effect of electronic photo enhancement on photographs. In most areas of the picture, the resolution is enhanced, but in some areas of the picture, the original picture becomes more corrupted after enhancement than before. In a rod drive, this means that, every once in while, we can expect to have long, burst errors and we also cannot necessarily expect the flux transition positions after compensation to be perfectly normally distributed within a detection interval (window).

The independent clock track helps the context-dependent postcompensation system since we may encounter areas on a disk with long defects. Conventional PLLs that derive timing information from the signal would have a hard time riding through the bad spots. This will be easier with a PLL that always maintains frequency lock.

A block diagram of the postcompensation system being proposed by ECC Technologies is shown in Figure 25.

Assume the detected transition positions are identified by leading or trailing edges of a sequence of digital pulses. The context tapped delay line is a snapshot of a targeted transitions context when the targeted transition (or pulse edge) is in the center of the delay. At that time, a time shift adjustment is fetched from a programmable channel characterization memory based on the context. It is possible, and may be more practical, to implement the channel characterization memory with simple digital logic which maps one set of bits into another set of bits. The time shift adjustment is loaded into a programmable delay line to subtract the time offset of the detected transition due to its context. This circuit will be practical because of the exponential growth in the density of digital ICs. (See ECC Technologies RAID & ECC Seminar material for a chart showing the exponential growth in number of transistors per chip.)

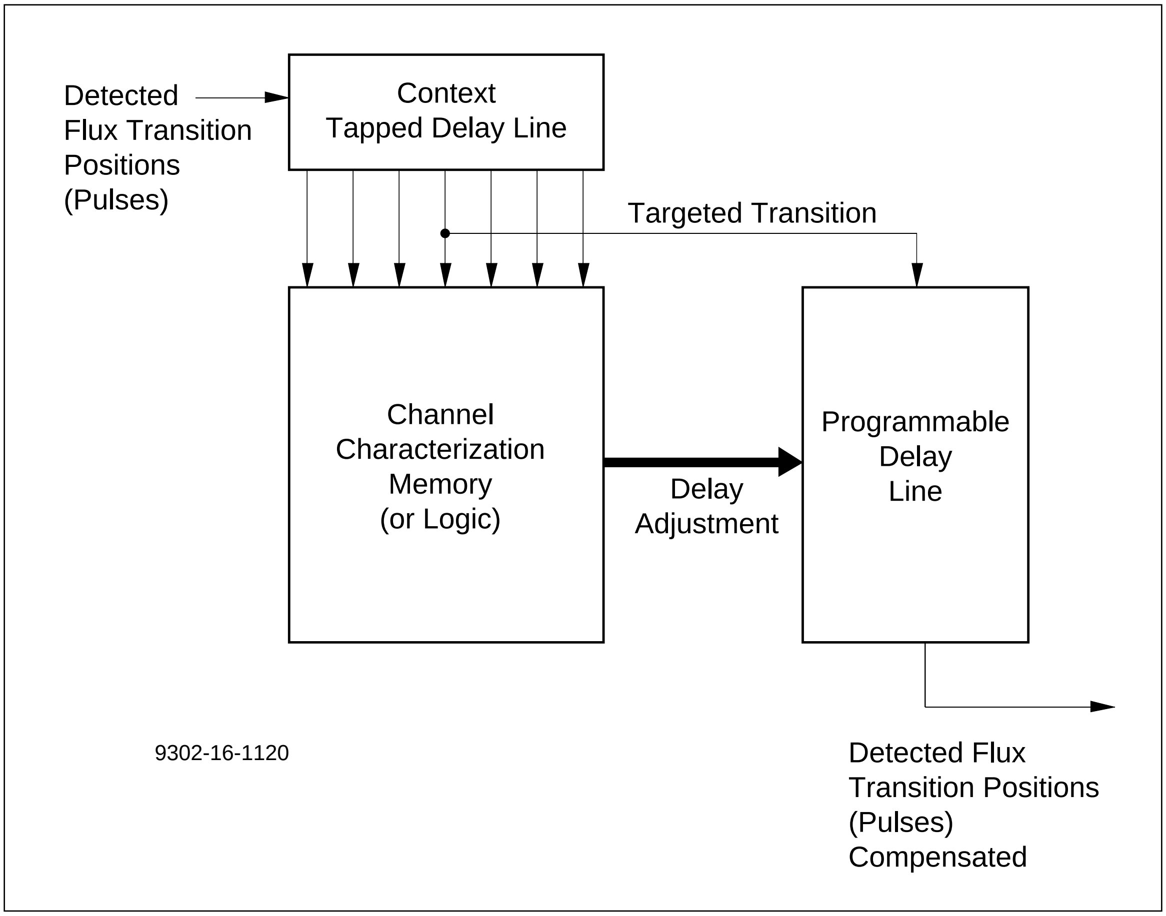


Figure 25 Block Diagram of Context-Dependent Postcompensation System