



Scsibench: User Level Emperical SCSI Disk Feature Extraction

Zoran Dimitrijevic and David Watson
CS290E Winter 2000



Introduction

- ❑ Problem: Design, implementation, and evaluation of a suite of programs that automatically determine the parameters of a SCSI disk
- ❑ Useful for disk request scheduling, and disk simulation

Definitions

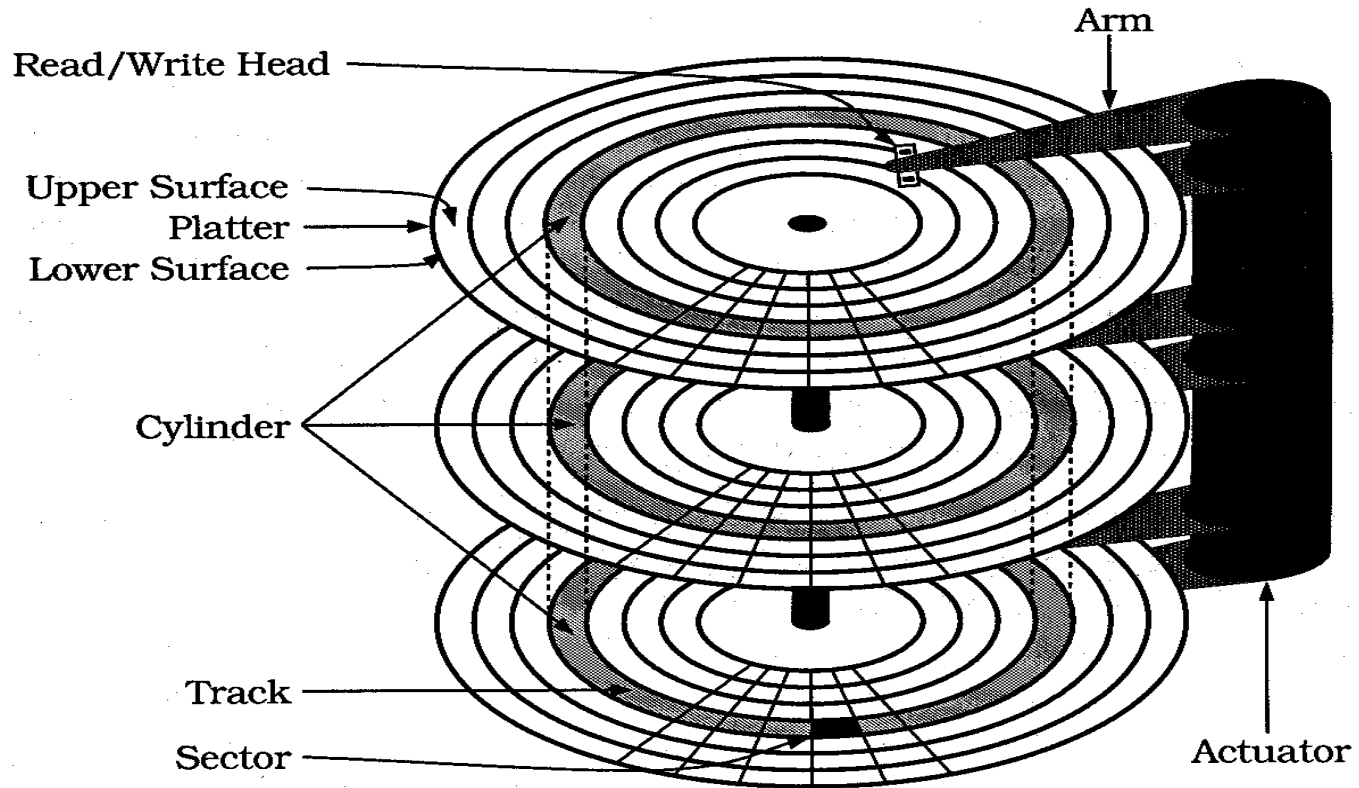
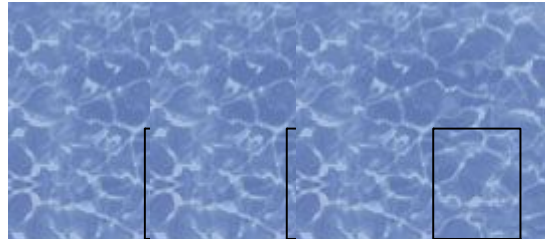


Figure 1: Disk drive terminology.

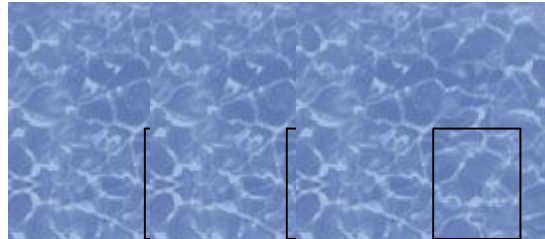
Definitions

- ❑ Cylinder Skew: Skew between last block on one cylinder and the first block on the next cylinder.
- ❑ Track Skew: Skew between last block on one track and the first block on the next track of the same cylinder.
- ❑ Cache: Divided into large fully associative blocks. Loaded with requested and/or prefetched data.
- ❑ Prefetch: Policy determines how much data is prefetched, and which cache segment is used.



Feature Extraction

- ❑ Disk Features: Physical Characteristics, Cache Layout and Prefetching Policies
- ❑ 2 Methods:
 - ❑ Inquiry: SCSI provides commands to ask for features
 - ❑ Emperical Extraction: Issue specific series of commands, and extrapolate from timing results



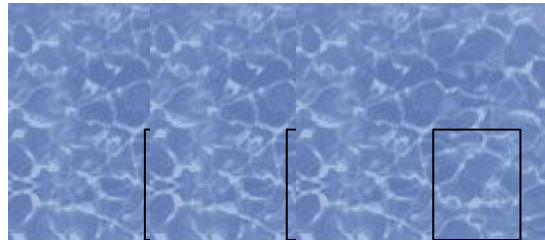
Motivation

- ❑ Inquiry isn't always accurate
 - ❑ Averages used for sectors/track, tracks/zone, and skews
- ❑ Many Inquiries are optional and not widely supported
- ❑ Many policies/settings not quantified in SCSI mode pages



Scsibench Overview

- ❑ Emperically measures:
 - ❑ Physical Characteristics:
 - ❑ Geometry:
 - ❑ Number of **Cylinders**
 - ❑ Number of **Zones**
 - ❑ **Logical to Physical** mapping (Cylinder, Head, Sector)
 - ❑ Rotation Speed (**RPM**)
 - ❑ **Head Switch** Time
 - ❑ Complete **Seek curves** (average/maximum)
 - ❑ **Transfer bandwidth** per zone



Scsibench Overview

- ❑ **Disk Cache:**
 - ❑ **Number of Cache Segments**
 - ❑ **Segment Size**
- ❑ **Prefetching:**
 - ❑ **Prefetching possible?**
 - ❑ **Prefetching Policy**
 - ❑ **Write Behind possible?**
- ❑ **Runs from user level, no kernel/driver modifications**

Scsibench Overview

- ❑ Interrogative:

- ❑ Physical Characteristics:

- ❑ Geometry:

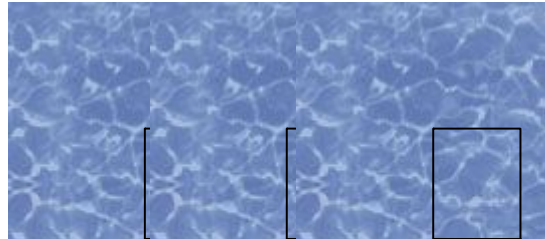
- ❑ Number of **Cylinders**

- ❑ Number of **Zones**

- ❑ **Logical to Physical** mapping (Cylinder, Head, Sector)

- ❑ Rotation Speed (**RPM**)

- ❑ Number of **Cache Segments**



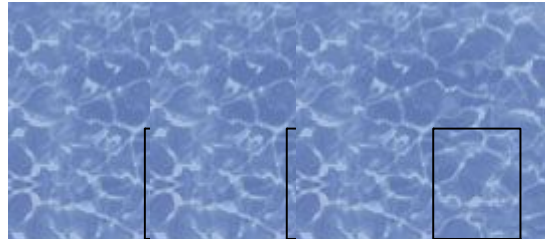
Implementation Details

- ❑ Implemented in Linux 2.2.x, using the SCSI Generic (sg) driver.
 - ❑ SCSI commands are built by the application, then written to the device. Responses are read back from the device.
 - ❑ Timing done using the cycle timer register on the Pentium.
- ❑ Emperical extraction uses only READ, WRITE, and CACHE DISABLE commands



Basic Technique

- ❑ Build a series of SCSI commands.
- ❑ For each command:
 - ' record start time
 - ' **write()** the command to the device
 - ' read() the result from the device
 - ' record end time
- ❑ Feature can be extrapolated from command completion times.

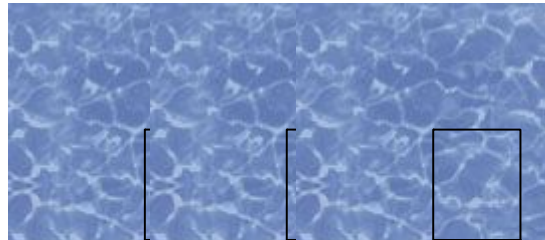


Basic Technique

❑ Example to find Rotation Rate:

(with read cache and prefetching OFF):

- ❑ <READ addr=X len=1 sector> time = 342 us
- ❑ <READ addr=X len=1 sector> time = 6000 us
- ❑ <READ addr=X len=1 sector> time = 6000 us
- ❑ $(1 \text{ min}) / (6000 \text{ us}) = 10000 \text{ RPM}$
- ❑ Disk must make 1 full rotation to re-read the same sector.



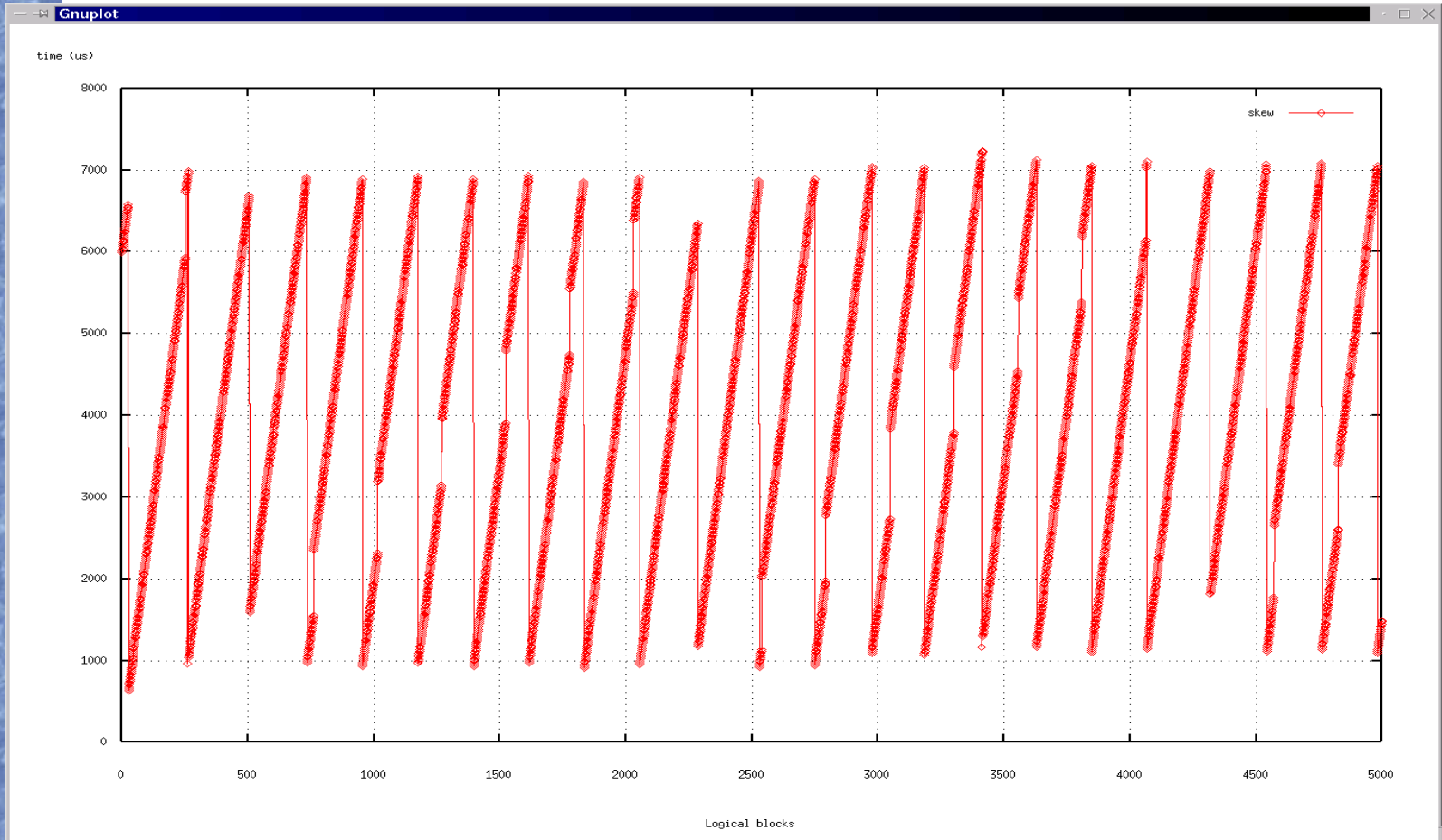
Rotation Rate

- $T1 = N1 * TR + T1SR + Thost1$
 $T2 = N2 * TR + T2SR + Thost2$
 $T2 - T1 = (N2 - N1) * TR + (Thost2 - Thost1)$
 $TR = ((T2 - T1) - \Delta Thost) / (N2 - N1)$
- Example:
Inquiry: $TR = 10045$
Emperical: $TR = 10045.488$

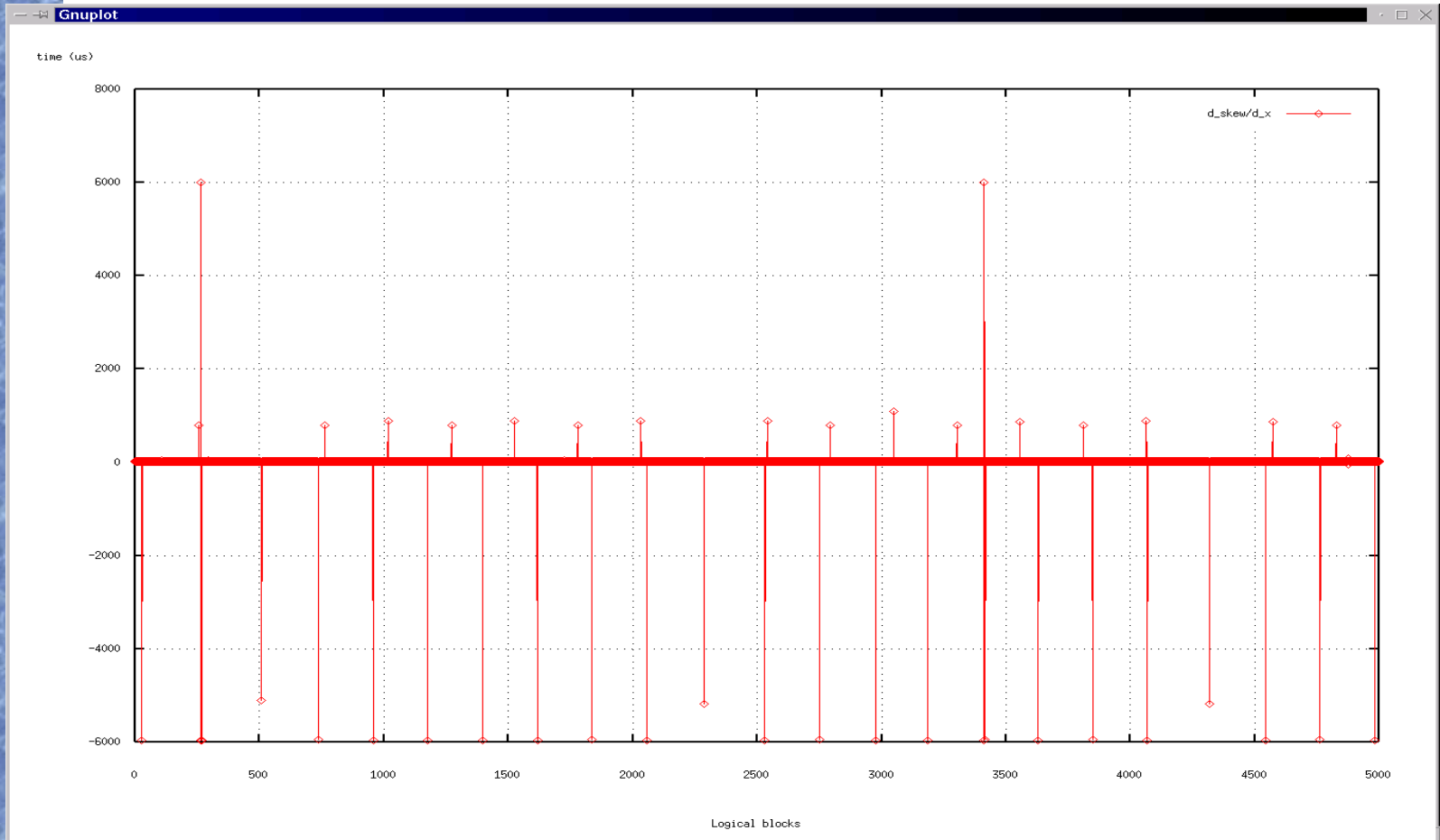
Head Switch and Settle

- ❑ $T1 = \text{MTBRC}(\text{write, read on next track})$
 $T2 = \text{MTBRC}(\text{write, read on same track})$
 $\text{THS} = T1 - T2$
- ❑ $T1 = \text{MTBRC}(\text{read, read on same track})$
 $T2 = \text{MTBRC}(\text{seek, read, write on same track})$
 $\text{TSETTLE} = T1 - T2$

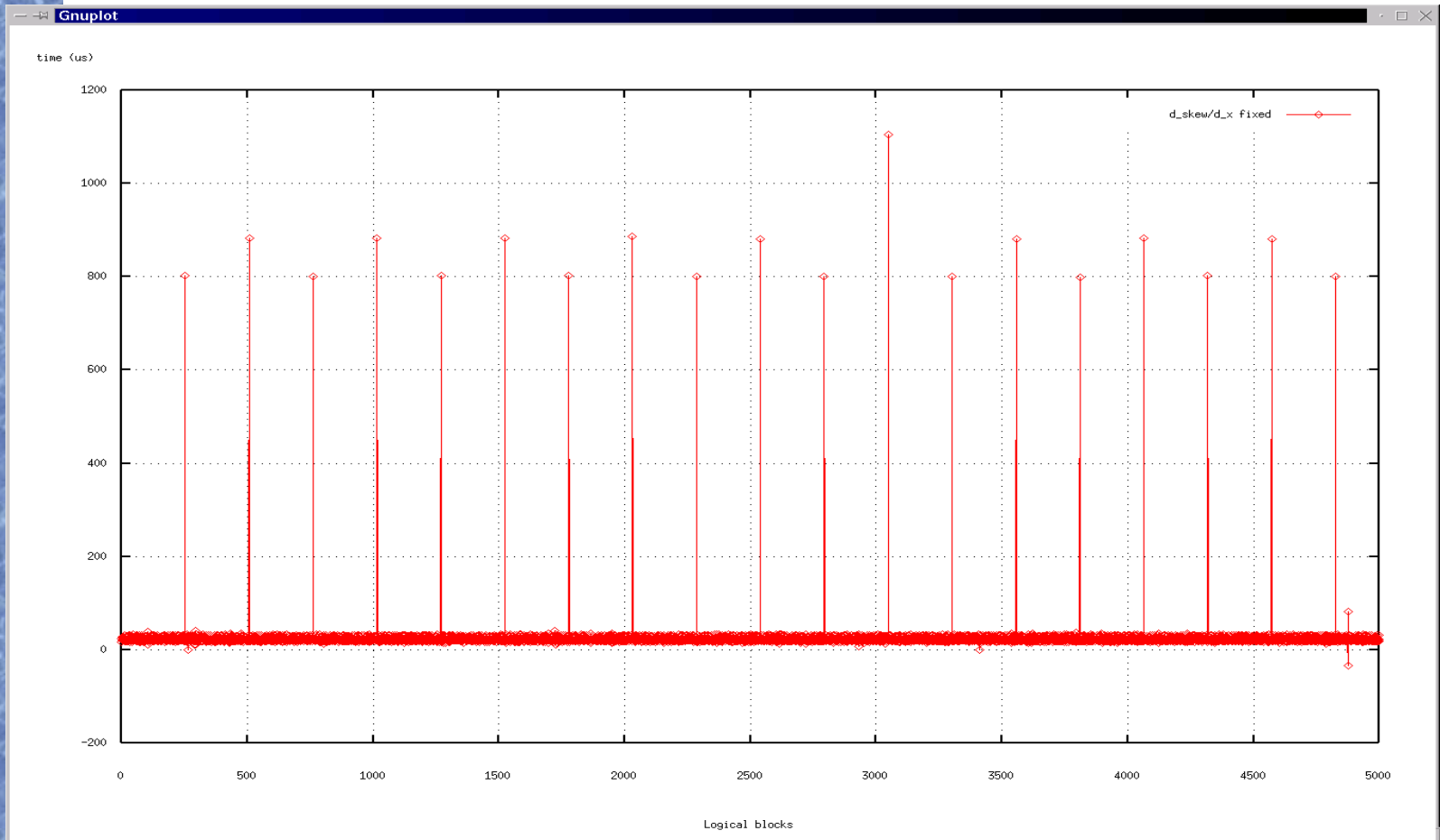
Emperical Mapping



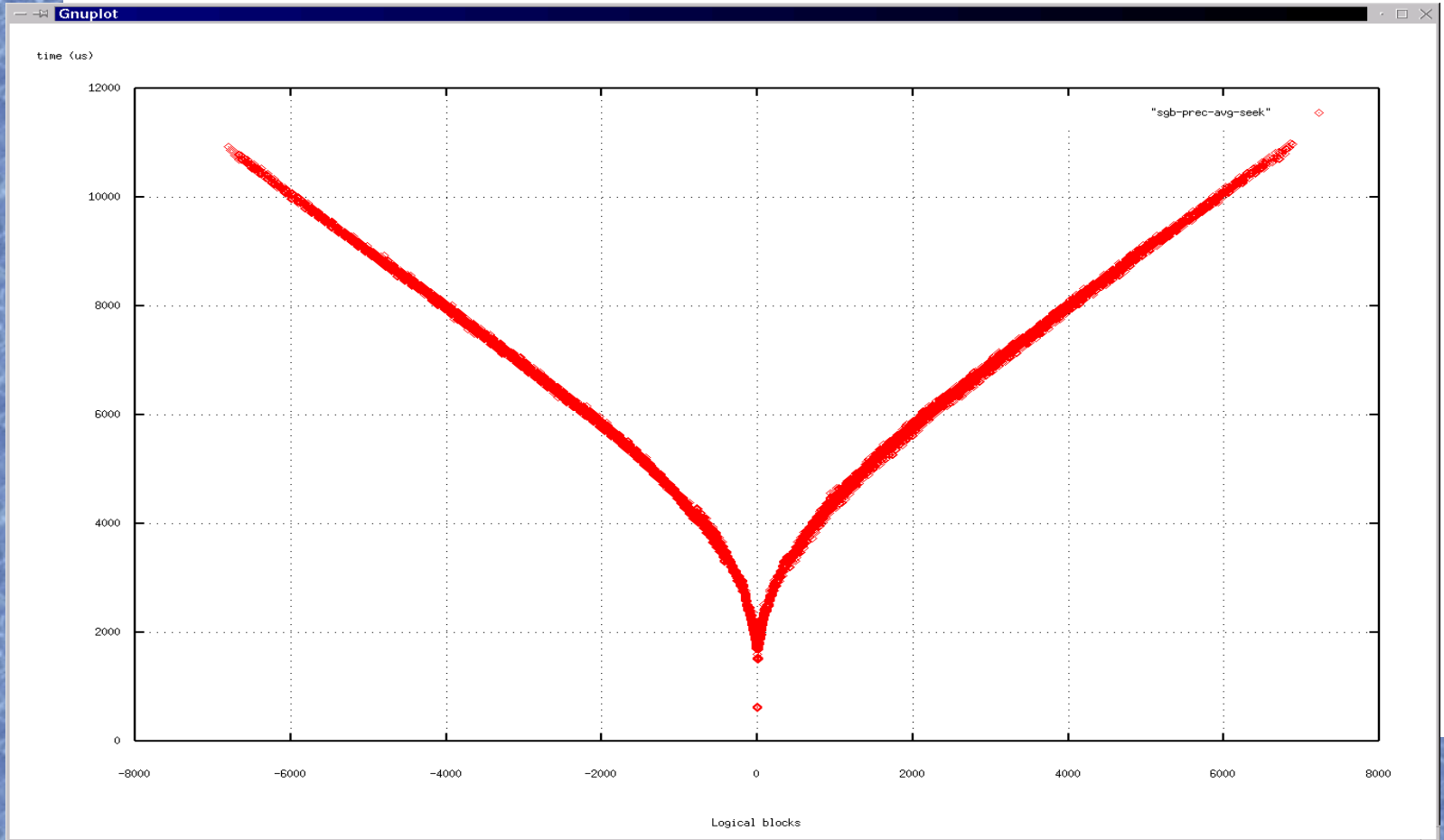
1st Derivative



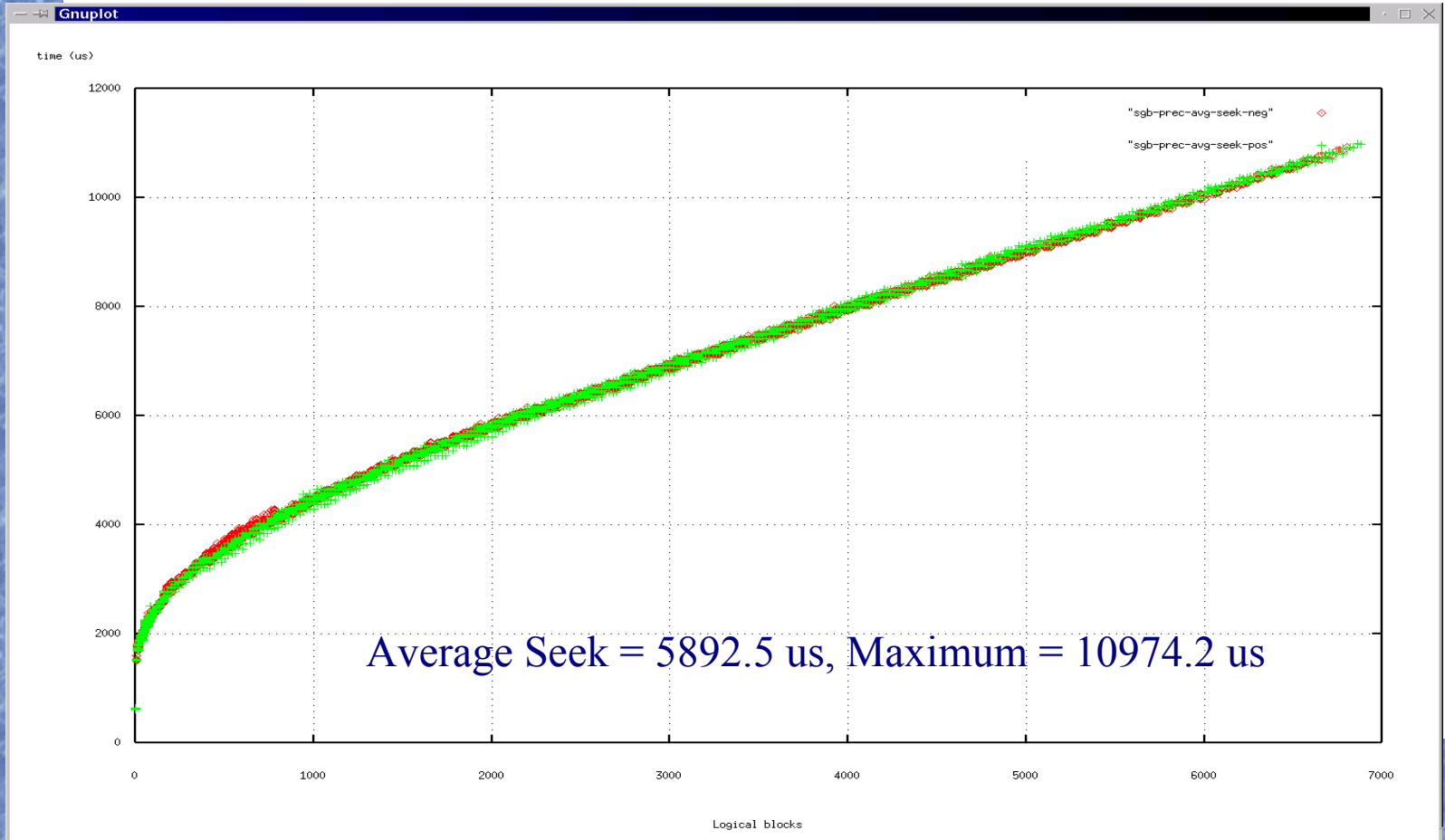
1st Derivative (Normalized)



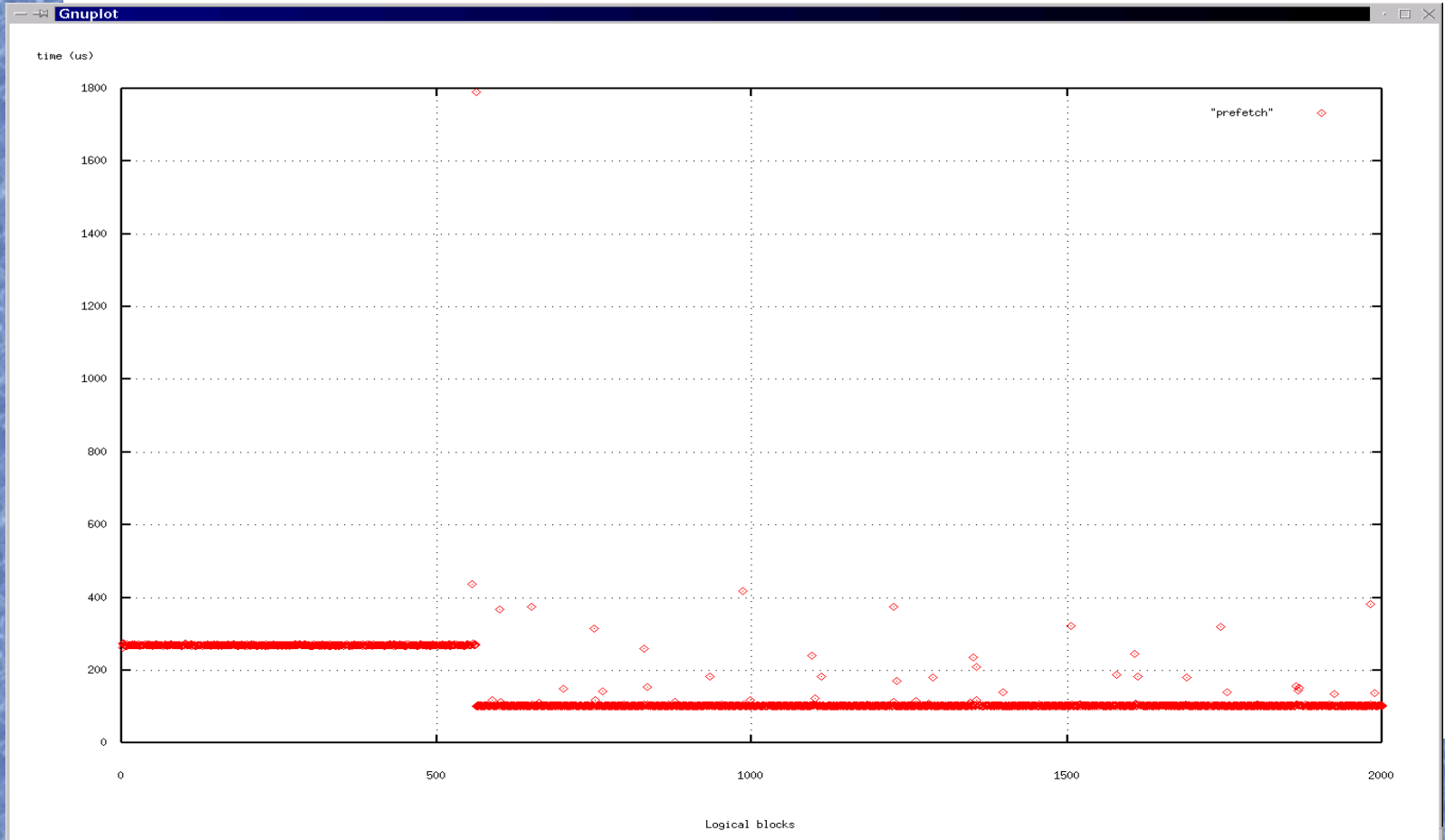
Seek Curves



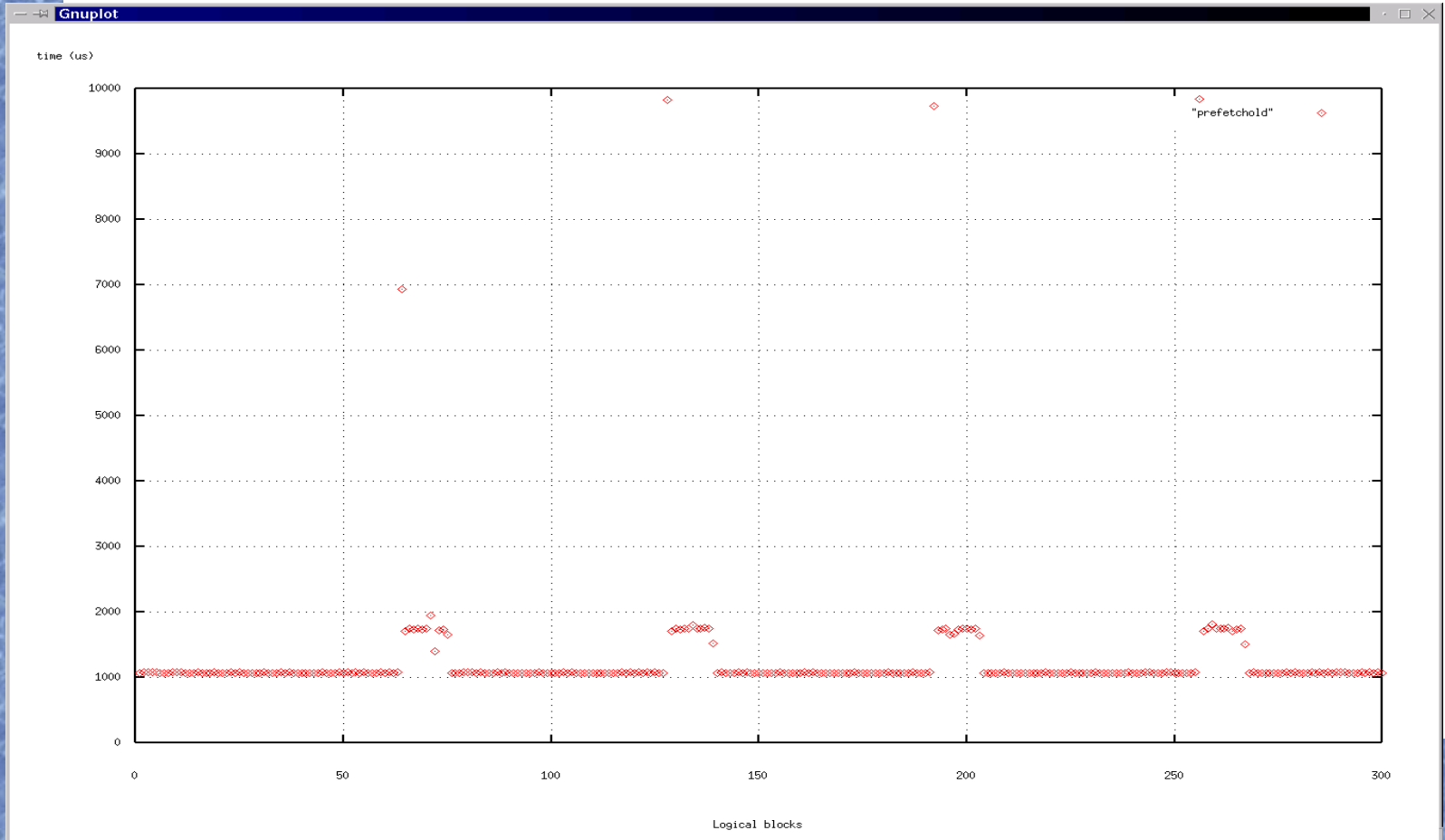
Seek Curves (Abs)



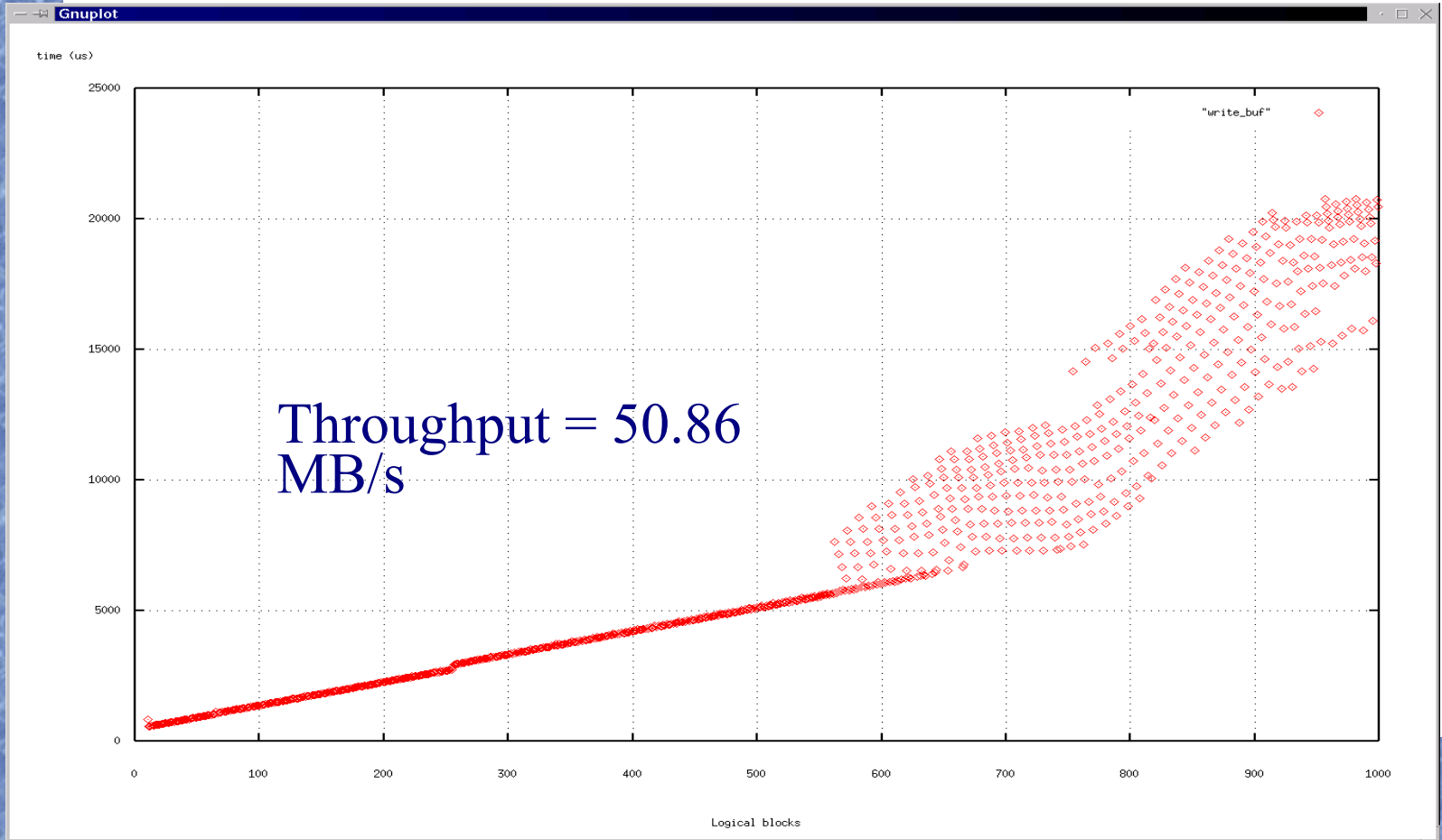
Prefetching



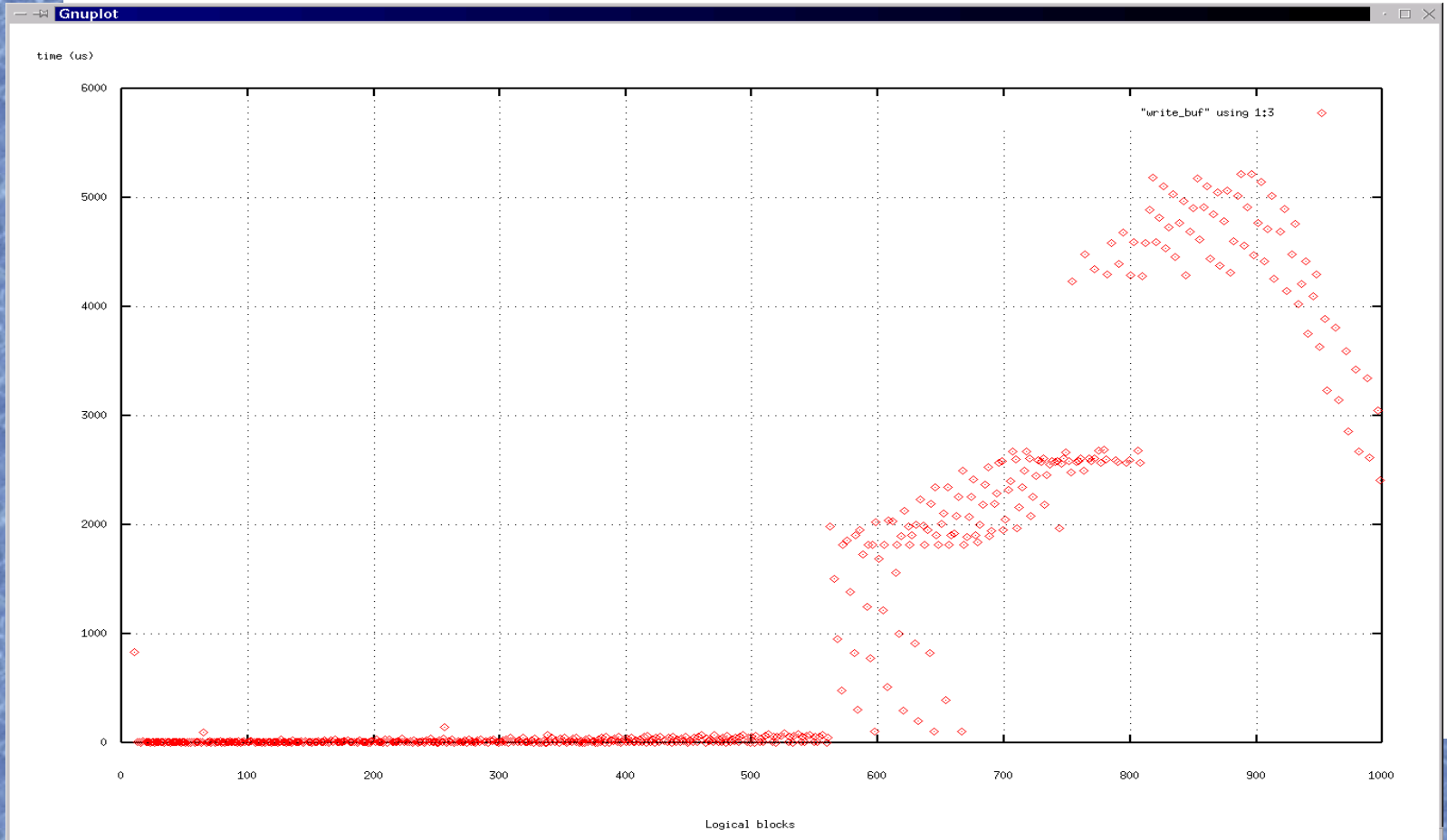
Prefetching (old disk)



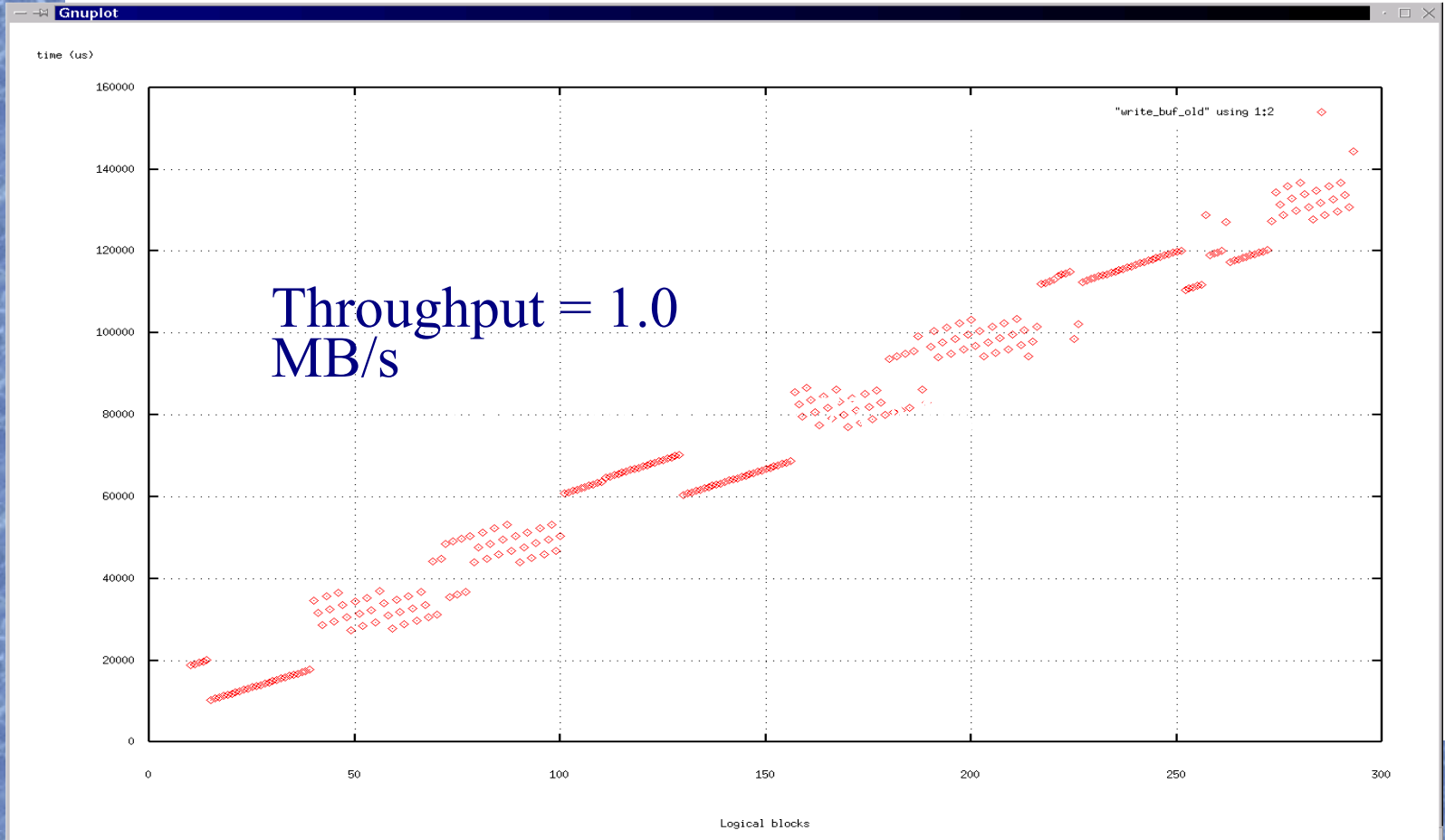
Write Buffer



1st Derivative



Write Buffer (Old Disk)



References

- ❑ Seagate SCSI Interface Product Manual, Volume 2
<http://www.seagate.com/support/disc/manuals/scsi/38479j.pdf>
- ❑ Notes on Linux's SG driver version 2.1.36
http://www.torque.net/sg/p/scsi-generic_long.txt
- ❑ ANSI SCSI-2 Working Draft Rev 10L (X3.131-199x)
<http://www.cs.ucsb.edu/~acha/courses/00/290e/scsi-2.txt.gz>
- ❑ The Programmer's Guide to SCSI, Brian Sawert
- ❑ The Linux SCSI programming HOWTO v1.5
<http://howto.tucows.com/LDP/HOWTO/SCSI-Programming-HOWTO.html>

