

LDPC Soft Decoding with Improved Performance in 1X-2X MLC and TLC NAND Flash-Based Solid State Drives

Lorenzo Zuolo, Cristian Zambelli, *Member, IEEE*, Alessia Marelli, Rino Micheloni, *Senior Member, IEEE*, and Piero Olivo

Abstract—The reliability of non-volatile NAND flash memories is reaching critical levels for traditional error detection and correction. Therefore, to ensure data trustworthiness in nowadays NAND flash-based Solid State Drives, it is essential to exploit powerful correction algorithms such as the Low Density Parity Check. However, the burdens of this approach materialize in a disk performance reduction. In this work a standard decoding approach is compared with an optimized solution exploiting hardware resources available in NAND flash chips. The simulation results on 2X, 1X and mid-1X MLC and TLC NAND flash-based Solid State Drives in terms of disk bandwidth, average latency, and Quality of Service favor the adoption of the presented solution in different host scenarios and realistic workloads. The proposed solution is particularly effective when high error correction interventions and read- or write-intensive workloads are considered.

Index Terms—Solid-State Drive, SSD, ECC, Low Density Parity Check, LDPC, Endurance, NAND Flash, MLC, TLC

I. INTRODUCTION

Solid State Drives (SSDs) are now the most effective solution for fast mass storage systems in cloud services and high performance computing [1]. One main SSDs' limitation is their reliability, which is dependent on the non-volatile NAND flash memories used as storage medium. These components, in fact, are subject to a progressive wear-out whose physical roots reside in the tunnel oxide degradation related to the mechanisms exploited for their program/erase. The aggressive technology scaling and the need for increasing memory capacity by storing more bits in a single cell (two bits Multi-Level Cells -MLC- or three bits Three-Level Cells -TLC- architectures) amplify the memories' wear-out impact on the SSD reliability [2], [3]. In fact, as long as the number of bits stored in a single cell increases, the width of the threshold voltage distribution associated to a logical stored content decreases. As a consequence, the control of the entire set of voltage distributions, which drift with the endurance (i.e., number of program/erase -P/E- cycles) and retention time, is becoming more and more complex. A direct indication of this phenomenon is an increase in the Raw Bit Error Rate (RBER)

Lorenzo Zuolo, Cristian Zambelli, Rino Micheloni and Piero Olivo are with the Dipartimento di Ingegneria, Università degli Studi di Ferrara, via G. Saragat,1 - 44122 Ferrara (Italy).

Alessia Marelli is with Microsemi, Via Torri Bianche 1, 20871 Vimercate (Italy).

Rino Micheloni is also with Microsemi, Via Torri Bianche 1, 20871 Vimercate (Italy).

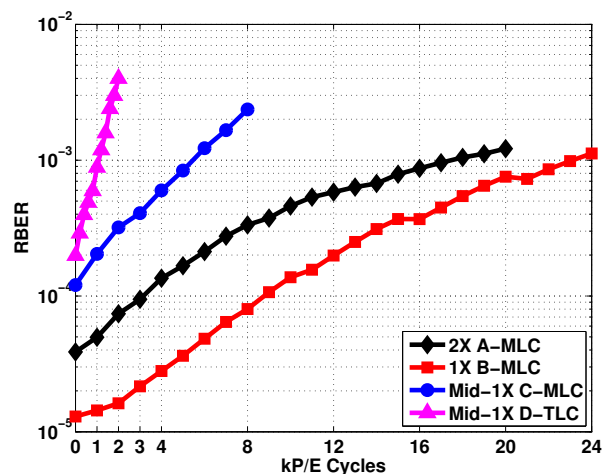


Fig. 1. Measured average RBER up to twice the rated endurance in 2X, 1X and mid-1X technology node MLC and TLC NAND flash memories as a function of P/E cycles.

in a NAND flash memory, that is the probability of having bits in error after a single read operation [4]. Such an increase translates into the inability to read correct data after a number of P/E operations or after long retention times. Fig. 1 shows the measured average RBER as a function of endurance for three MLC and one TLC NAND flash memories manufactured in 2X, 1X, and mid-1X technology nodes as described in Table. I. As it can be seen, as the number of P/E cycles increases, the error rate quickly grows up. In addition, either by scaling from a 2X to a mid-1X node or switching from a MLC to a TLC storage paradigm, the RBER increases significantly.

To broaden NAND flash reliability figures and, consequently, data trustworthiness over the whole SSDs' lifetime, the use of sophisticated Error Correction Codes (ECC) is essential. This requirement is tightly coupled with the percentage of uncorrectable pages in a NAND flash memory that are pages which, if read, return a number of errors greater than the ECC's correction limit. This latter value represents a quality metric of the whole SSD's reliability because as soon as it is reached, NAND flash memories and therefore the disk are considered as failed [5]. Table. II shows the endurance measured for the 4 considered memories when a multi-threaded BCH decoder able to correct up to 100 errors in a 4320 Bytes codeword is used [6]. To overcome these limitations with the aim of moving the disk failure point as far

TABLE I
MAIN CHARACTERISTICS OF TESTED NAND FLASH MEMORIES.

Sample	A-MLC	B-MLC	C-MLC	D-TLC
Memory type	Consumer	Enterprise	Enterprise*	Enterprise*
Rated endurance	9 k P/E	12 k P/E	4 k P/E	0.9 k P/E
Measured Average Read Time [μ s]	68	40	70	86
Measured Average Program Time [μ s]	1400	2000	2500	2300
Program mode	dual plane	dual plane	dual plane	dual plane
Page size [Bytes]**	16384	16384	16384	16384
Technology node	2X	1X	Mid-1X	Mid-1X

*Early samples

**w/o spare area

TABLE II
MEASURED ENDURANCE SUSTAINED WITH A BCH ECC ENGINE.

Sample	Measured endurance
A-MLC	6 kP/E
B-MLC	19 kP/E
C-MLC	5 kP/E
D-TLC	1 kP/E

as possible, a more powerful correction code must be adopted.

Due to their superior error correction capabilities, Low Density Parity Check (LDPC) codes now represent a forced choice for SSDs [7], [8]. Conventional LDPC decoders, if properly designed, can sustain a NAND flash RBER up to 10^{-2} [8], [9], [10], [11]. The LDPC correction engine usually leverages on two sequential correction approaches: *i*) the hard decision (HD) which corrects errors by means of a single read operation of the selected memory page; *ii*) a sequence of soft-level decisions (SD) which perform, with considerably higher latency, a fine-grained multiple-read sensing operation that allows error correction by combining the multiple-read data with the original HD.

As summarized in Fig. 2, besides the HD whose data are stored in a buffer inside the LDPC decoder as a reference, each soft-level requires two page read operations with two different read references and two data transfers to the ECC engine. The algorithm continues this process until the page is correctly read or the maximum number n of soft-levels is reached and the page is marked as uncorrectable. The overall n -level SD algorithm requires $2n$ page reads and $2n$ data transfers operations. This serial approach is used mainly because high code-rates [12] are adopted to exploit the full SSD capacity and hence HD has the same limitations of BCH codes in terms of RBER [11]. Therefore as soon as this strategy fails to correct data, it is requested the intervention of the SD, with a higher correction range. However, in [11] it has been shown that, as soon as the HD approach starts to fail, there is an overhead both in terms of increased SSD power consumption and overall SSD latency since additional read operations are requested on NAND flash with respect to the HD approach.

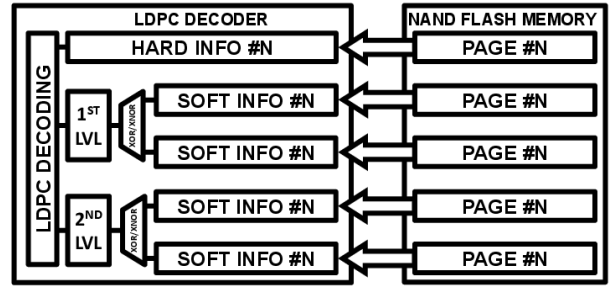


Fig. 2. Standard LDPC decoding (HD + two-level-SD). Besides the HD whose data are stored inside the LDPC decoder, each soft-level requires two extra read operations and two data transfer operations.

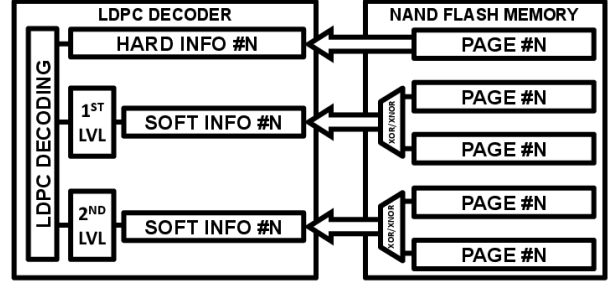


Fig. 3. LDPC decoding (HD + two-level-NASD). Besides the HD whose data are stored inside the LDPC decoder, each soft-level requires two extra read operations and only one data transfer since read data are conventionally combined before data transfer.

An alternative LDPC correction approach that limits the drawbacks of the SD has been presented in [13]. The assumption of this methodology, named NAND-Assisted Soft Decision (NASD), is that data for ECC engine are produced by the NAND flash memory itself, which internally reads the target page twice for each soft-level. Then, read data are opportunely combined and only one transfer to the ECC is performed for each soft-level, as shown in Fig. 3, thus reducing the NAND flash I/O bus use. The NASD advantages become more pronounced when the impact on the command scheduling by the halved number of data transfers is taken into account.

In this paper we apply the NASD technique on 2X, 1X, and mid-1X MLC and TLC NAND flash-based SSD architectures to:

- show how NASD, thanks to a reduced number of data transfers and to the consequent impact on command scheduling, modifies significantly the SSD figures of merit: bandwidth, average latency, NAND flash I/O bus use, and Quality of Service (QoS) that is the ability of keeping a sustained performance over time within a defined threshold [14], [15], [16];
- compare the SSD performance at system level obtained exploiting the standard HD+SD and the HD+NASD LDPC. The analysis have been performed on two different host architectures: a consumer PC and an enterprise workstation;
- show, for the two host architectures, how NASD outperforms the traditional SD approach when synthetic 100% read and different realistic workloads such as MSN,

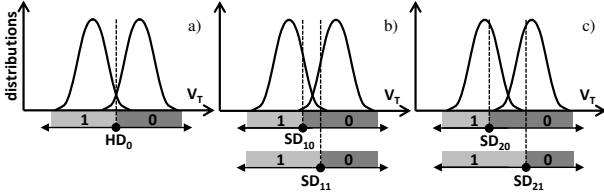


Fig. 4. NAND flash read references used in the two levels LDPC sensing scheme. A memory page is read by setting the read voltage at HD_0 and determining, for each bit, whether $V_T < HD_0$ or $V_T > HD_0$ (a). If the ECC engine is not able to correct possible read errors, the soft decision algorithm starts and the page is read twice by moving the read references around HD_0 , at SD_{10} and SD_{11} (b). If the page is still marked as uncorrectable, the page is read again with the SD_{20} and SD_{21} references (c).

Financial, and Exchange [17] are considered;

The system performance have been evaluated by using the SSDEplorer co-simulation framework [18], [19].

II. SOFT DECISION VS NAND-ASSISTED SOFT DECISION

NAND flash memories are read page-wise by using a defined read reference, hereafter denoted as HD_0 . Cells are read as 1 or 0 depending on their threshold voltage V_T with respect to HD_0 (see Fig. 4a). If during the ECC decoding phase the page is evaluated as uncorrectable, the LDPC decoding algorithm can be retried with the SD. To accomplish this second step, more information about the actual position of the NAND flash threshold voltage distributions must be collected. Basically, the algorithm moves sequentially the internal read references to SD_{10} and SD_{11} (Fig. 4b) thus reading the page twice and storing the two data content in two page registers inside a page buffer. Data from the page buffer are transferred byte-wise from the flash memory to the LDPC decoder and then are analyzed with those previously read with HD_0 . This step is possible because during the whole SD process the data read with the HD_0 are buffered inside the LDPC decoder and are used as a reference. If the decoding process still fails, a second iteration is performed by moving the read references to SD_{20} and SD_{21} and comparing the new read data with the HD as shown in Fig. 4c. The algorithm continues this process until the page is correctly read or the maximum number of soft-levels is reached and the page is marked as uncorrectable.

Table. III summarizes the number of operations performed by both algorithms. As it can be seen, NASD is able to halve the number of page transferred from the NAND flash memory to the ECC. As a consequence, the overall soft decision process is shortened and hence, the SSD performance are improved. To understand the effective NASD efficiency, it must be taken into account that read operations are temporally separated from the successive data transfer operations. Fig. 5 sketches the commands queue for NAND flash dies sharing the same I/O bus, the corresponding data bus allocation, and the ECC engine activity. After a HD_0 read, the SSD controller can send other read or write commands to the same NAND flash die or to other dies. When the ECC engine communicates the read failure to the controller, this latter stores the data related to the HD and schedules the additional SD_{10} and SD_{11} reads. In the SD approach the two read data are transferred

TABLE III
READ AND DATA TRANSFER OPERATIONS IN SD AND NASD APPROACHES.

LDPC	One soft-level	Two Soft-levels	#n soft-levels
SD	HD + 2 page read + 2 data transfer	HD + 4 page read + 4 data transfer	HD + #2n page read + #2n page transfer
NASD	HD + 2 page read + 1 data transfer	HD + 4 page read + 2 data transfer	HD + #2n page read + #n page transfer

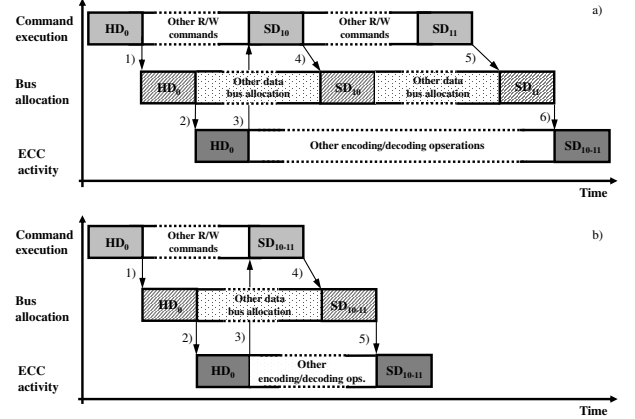


Fig. 5. Time sketch, for a cluster of NAND flash dies sharing the same data bus, of the command queue, of the data bus allocation, and of the ECC engine activity. Numbers enlighten the events sequence during a single soft-level decision operation. Case a) and b) refer to the SD and NASD approach, respectively.

separately when the I/O bus is available, with the risk that between the SD_{10} and SD_{11} transfer the bus is contended by other data transfers to/from other NAND flash dies (see Fig. 5a). In the NASD approach, on the contrary, since SD_{10} and SD_{11} read data are combined in a single data transfer, the consequent soft decision operation can start in advance with respect to the SD case (see Fig. 5b). The advantages, that become more pronounced when additional soft-levels are considered, depend on the considered workload, as shows in Section III-B. Moreover, since the number of data transfers between the memory and the ECC are reduced, NAND flash memory I/O bus accesses are reduced as well. This I/O bus use reduction impacts the SSD dynamic power consumption.

The main component exploited by NASD is the NAND flash page buffer which is used to store data for each soft-level operation. In present NAND flash chips, this buffer is composed by two registers used especially for read cache and read retry operations [20], [21], [22], [23]. It becomes clear that the NASD implementation does not require any other register inside the memory and it can be performed by a simple 8-Bit combinational logic placed between the internal NAND flash page buffer and the I/O interface. In fact, the two read operations performed by NASD can be easily stored into the existing registers of the page buffer and a simple block composed by 8 XORs (or 8 XNORs) acting as a combinational circuitry is sufficient. Since the I/O interface limits the parallelism to 8-bits, the logic combination between the pages stored inside the two registers can be performed on-

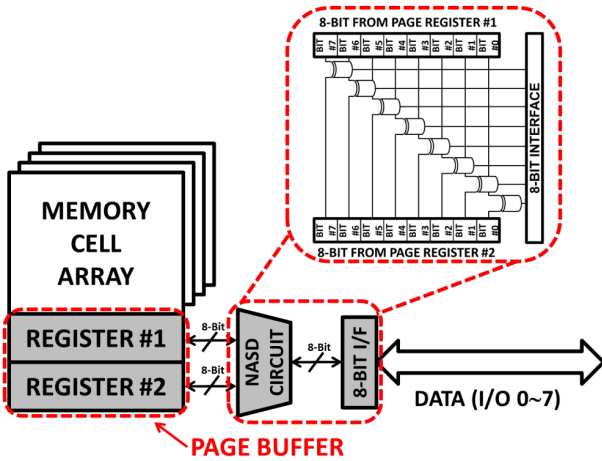


Fig. 6. NASD combinational circuitry. Just 8 logic XOR gates (or XNOR) have to be added before the 8-bit I/O interface. Data read from the NAND flash array are temporarily stored in Register #1 and Register #2. After that they are byte-wise combined by the NASD circuit and transferred to the 8-bits I/O bus.

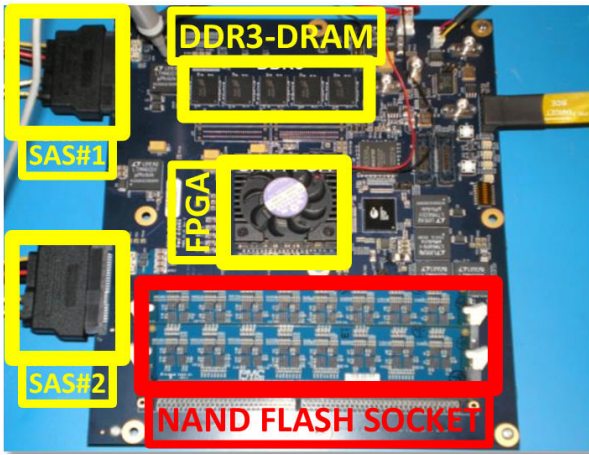


Fig. 7. Characterization board used to stress the tested NAND flash chips.

the-fly in a byte-wise fashion during the data transfer phase (see Fig. 6). Regardless of the internal NAND architecture, just a single combinational logic can be integrated in a single NAND chip. As a consequence, the NASD implementation inside a NAND flash memory becomes a easy task which does not impact neither chip area nor power consumption.

III. EXPERIMENTAL SETUP AND RESULTS

Results have been collected by means of: *i*) a dedicated NAND flash memory characterization system which collects RBERs and statistics on uncorrectable pages; *ii*) a hardware implementation of the LDPC code which computes real decoding latencies; *iii*) a co-simulation SSD framework able to produce bandwidth, latency of a target disk architecture starting from previously collected reliability data and ECC statistics [18], [19].

Fig. 7 shows the test equipment exploited for memories characterization. It is composed by a programmable FPGA, a DRAM buffer for temporary data storage and a dedicated socket for NAND flash memory interfacing. Each tested device

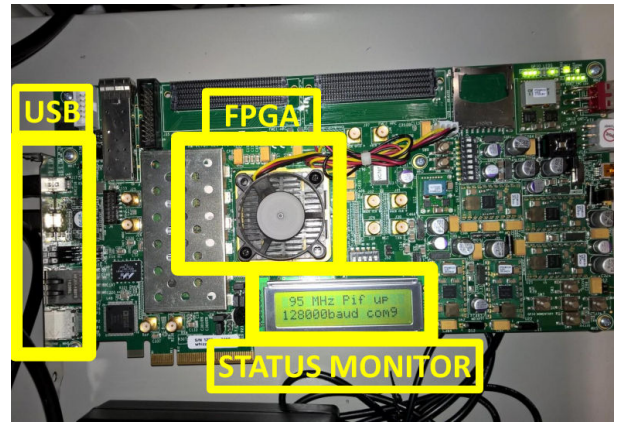


Fig. 8. LDPC characterization board for ECCs decoding and encoding latencies evaluation.

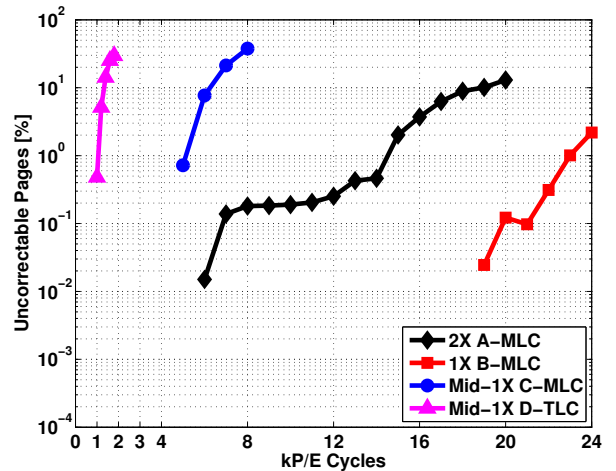


Fig. 9. Measured percentage of uncorrectable pages as a function of P/E cycles when only HD-LDPC is used with a capability to correct up to 100 bits in error in a 4320 Bytes codeword.

has been sequentially stressed with random data patterns. For testing purposes, each NAND flash memory has been stressed with a number of P/E cycles higher than their rated endurance (Table. I).

Fig. 8 shows the LDPC characterization setup. The board has been configured to generate random data patterns emulating different RBER values from a NAND flash. The codeword is computed and decoded by the LDPC board, whereas an external PC gathers encoding and decoding latencies to be further exploited by the co-simulation SSD framework. The HD correction capability of the LDPC engine has been set with the same correction strength used by the BCH code described in the Introduction (i.e., up to 100 errors in a 4320 Bytes codeword).

Fig. 9 shows the percentage of uncorrectable pages when only a HD-LDPC approach is used. As it can be seen, mid-1X MLC and mid-1X TLC memories show a high HD fail rate so that SD would be constantly required. As a consequence, NASD technique advantages are evident resulting in a higher SSD read bandwidth, an improved QoS and a lower average read latency. On the contrary, 1X-MLC and 2X-

TABLE IV
SIMULATED SSD ARCHITECTURE.

Parameter	Configuration
Channels	8
Dies per channel	8
Die capacity	128 Gb
SSD capacity	512 GByte
Host interface	PCI-Express Gen2x8 [24]
Host protocol	NVM-express 1.1 [25]

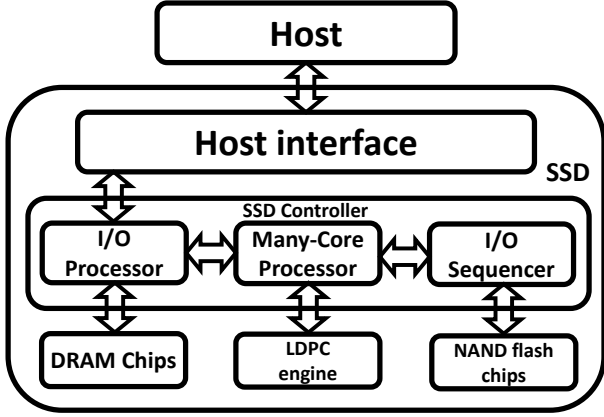


Fig. 10. Simulated SSD architecture.

MLC memories show a low percentage of uncorrectable pages which grows up only in proximity of the rated endurance. In these cases, error correction is less required and hence NASD advantages are present yet barely perceivable. It must be highlighted that two soft-levels were sufficient to correctly read all the tested memories up to twice the rated endurance, for both SD and NASD approaches.

The simulated SSD architecture is summarized in Table IV. Fig. 10 shows the main building blocks of the SSD. Besides the standard I/O processor exploited for the host-interface address fetch phase and the many-core processor on which the operations' scheduler is executed, there is also an I/O processor acting as a read/write dies sequencer. In order to fully exploit the internal parallelism offered by the SSD, host random addresses which could cause die collisions (i.e., requests for a die already scheduled) are parsed and sequentially issued to NAND flash chips. In such a way, even if random commands are sent by the host, only sequential patterns are processed by NAND flash memories hence maximizing the throughput. To achieve accurate simulation results, command scheduling phenomena such as queuing and pipelining have been considered.

All data have been collected simulating two different host platforms (Table. V). The first one is a consumer system which does not exploit the full SSD architecture (able to sustain 450 kIOPS) since I/O requests settle around 200 kIOPS. As a consequence, all internal error recovery techniques which exploit additional read operations produced by the ECC for the soft decoding step are partially hidden by the SSD's architecture which masks all non-user reads. The second one is an enterprise workstation designed to serve hundreds of parallel processes which requests up to 600 kIOPS. In this

TABLE V
TESTED HOST SYSTEM CONFIGURATIONS

	Consumer	Enterprise [26]
Host Processor	Intel-Core i5-4570	Intel-Xeon e5-2630
Processor clock	3.2 GHz	2.3 GHz
#N Cores	4	24
DRAM size	12 GByte	16 GByte
Workload generator [27]	fio 2.1.10	fio 2.1.10
Avg. I/O submission time	3.5 μ s	0.5 μ s
Host queue depth [25]	64	256
Host kIOPS (requested)	\approx 200	\approx 600
SSD kIOPS (sustained)	\approx 450	\approx 450

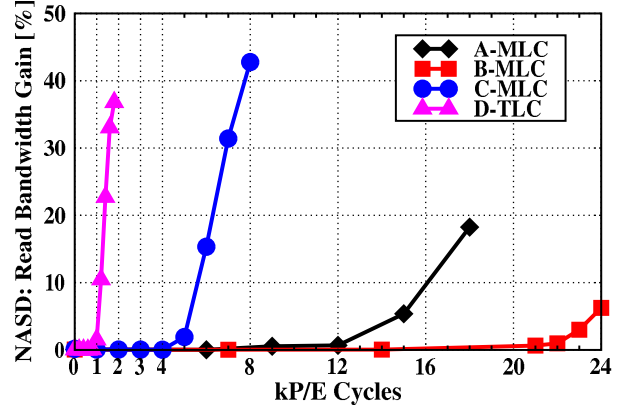


Fig. 11. SSD read bandwidth gain achieved by NASD with respect to two soft-levels SD as a function of the memory endurance for the 4 considered memory types and the Enterprise host.

case the disk performance cannot match this specification so that any further read produced by any error recovery technique will burden on the final SSD's performance. Thanks to these two different test-cases it has been possible to test the NASD effectiveness over standard SD when disk resources such as NAND-flash I/O buses are partially or completely allocated for user operations.

Results presented in Section III-A refer to an enterprise host and a 100% 4 kB random read workload which represents the most challenging situation for the SSD performance characterization. In fact, when mixed read/write workloads are considered, since the DRAM chip in the SSD caches all the write operations, the measured average latency and bandwidth figures of the disk do not reflect the actual SSD behavior. Section III-B will extend the discussion to realistic workloads for both hosts.

A. 100% random read workload - Enterprise host

Fig. 11 shows the SSD's read bandwidth gains achieved by the NASD approach with respect to the SD, as a function of the memory endurance. The simulations have been performed considering the 4 different memories as SSD's storage medium. Bandwidth (IOPS) has been calculated as the average number of read commands completed in a second. As it can be seen, for all the considered memories the NASD technique provides a significant gain. NASD advantages are more pronounced when large number of uncorrectable pages, triggering a massive ECC intervention, are detected.

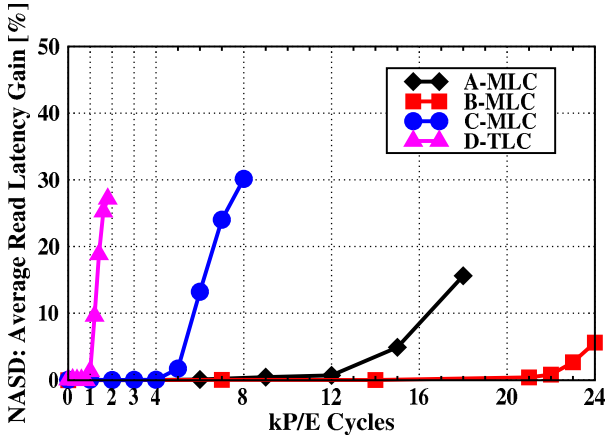


Fig. 12. SSD average read latency gain achieved by NASD with respect to two soft-levels SD as a function of the memory endurance for the 4 considered memory types and the Enterprise host.

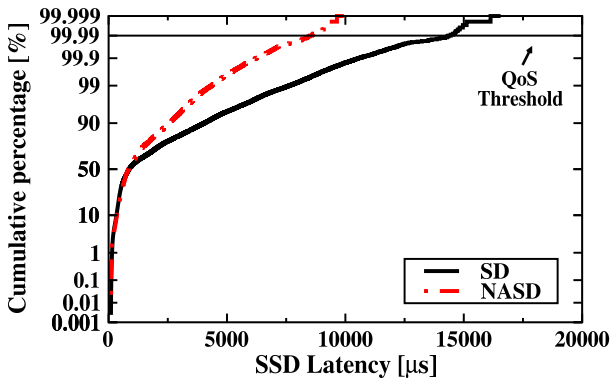


Fig. 13. Cumulative percentage on a normal probability paper of the SSD latency calculated at twice the rated endurance when a D-TLC sample is used and both SD and NASD are considered. The QoS threshold is calculated as the 99.99 percentile of the cumulative distribution [14].

Fig. 12 shows the average read latency gains achieved by NASD with respect to SD as a function of memory endurance. Latency has been calculated as the average time elapsed between a read command submission and its completion. All results concerning average latency reflect those obtained for bandwidth (Fig. 11).

Fig. 13 shows the SSD's cumulative latency distributions calculated at twice the rated endurance for the D-TLC sample and both SD and NASD approaches. From these data it is possible to extract the SSD's QoS defined as the 99.99 percentile of the cumulative latency distribution [14]. QoS represents the predictability of low latency and consistency of high bandwidth while servicing a defined workload and it can be considered as the key metric to assess the SSD's performance in a worst-case scenario. Fig. 14 shows the calculated QoS at twice the rated endurance for all the considered memories and for both the SD and NASD approaches.

B. Realistic workloads - Enterprise and Consumer hosts

Since the NASD advantages are tightly coupled to the RBER showed by the NAND flash memories and to the command pattern, simulations have been also performed con-

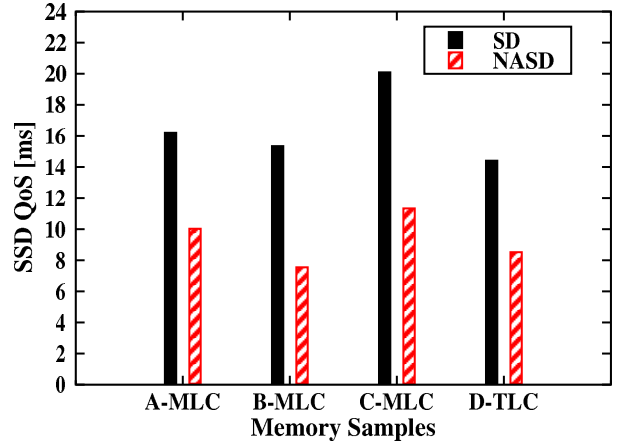


Fig. 14. Calculated QoS at twice the rated endurance for the 4 considered memory types and the Enterprise host.

TABLE VI
WORKLOADS CHARACTERISTICS

Workload	Write ratio [%]	Write amplification factor
MSN	96	1
Financial	81	1.32
Exchange	46	1.94

sidering three realistic workloads [17], as detailed in Table VI. *Write ratio* represents the percentage of write commands in the command sequence, whereas *write amplification factor* denotes the number of additional writes produced by the SSD firmware for each single host write [28].

In the MLC and TLC architectures the write throughput is smaller than the read throughput (see Table I). In fact, to lower the RBER retrieved during read operations, sophisticated but long program algorithms are used [2], [29]. To deal with this bandwidth mismatch, it is usual to leverage multi-plane program commands which allow writing, on the same memory die, two or more pages in the time-frame of a single page program. This approach, on the one hand allows maximizing the program throughput towards the NAND flash dies, on the other hand, however, it severely impacts the I/O bus transfer time. In fact, for each program operation two or more 16 kBytes pages have to be transferred from the controller to the target memory die thus making the I/O bus busy for long times. In the NAND flash memories considered in this work write operations are performed in a dual-plane mode (see Table I) Therefore, before scheduling the actual program operation on a memory die, a chunk of 32 kBytes has to be moved from the SSD controller to the NAND flash die. As a consequence, since 4 kBytes chunks are read by the host during a read operation, it is clear that when programs are scheduled, the I/O bus is busy for a time which is 8x longer than a read. In light of these considerations and taking into account the scheduling effects shown in Fig. 5, it is clear that NASD will show better results either in extremely write intensive workloads (i.e., MSN) or in read intensive workloads. In the former case the probability of having a long write transfer between the two read operations required by the standard SD technique is high, whereas in the latter case other read operations can

TABLE VII
BANDWIDTH (IN KIOPS FOR SD AND IN % OF GAIN FOR NASD VS SD) @ TWICE THE RATED ENDURANCE FOR BOTH THE CONSUMER AND THE ENTERPRISE HOST

Workload	Consumer Host								Enterprise Host							
	A-MLC		B-MLC		C-MLC		D-TLC		A-MLC		B-MLC		C-MLC		D-TLC	
	SD	NASD	SD	NASD	SD	NASD	SD	NASD	SD	NASD	SD	NASD	SD	NASD	SD	NASD
MSN	143	4.78	147	4.17	142	5.26	141	5.23	142	4.14	148	4.79	142	5.94	142	5.72
Financial	135	1.45	142	0.54	101	3.68	104	4.36	143	1.88	148	0.45	119	3.67	118	4.19
Exchange	143	2.25	151	0.60	94	4.95	97	5.40	171	2.47	187	0.44	126	5.28	127	6.0
100% read	204	0.03	204	0.03	140	24.41	127	24.85	299	18.24	402	6.24	152	42.77	156	36.80

TABLE VIII
AVERAGE LATENCY (IN μs FOR SD AND IN % OF GAIN FOR NASD VS SD) @ TWICE THE RATED ENDURANCE FOR BOTH THE CONSUMER AND THE ENTERPRISE HOST

Workload	Consumer Host								Enterprise Host							
	A-MLC		B-MLC		C-MLC		D-TLC		A-MLC		B-MLC		C-MLC		D-TLC	
	SD	NASD	SD	NASD	SD	NASD	SD	NASD	SD	NASD	SD	NASD	SD	NASD	SD	NASD
MSN	373	2.24	343	0.11	393	5.46	396	3.57	1485	0.57	1468	0.16	1553	1.78	1554	1.92
Financial	467	1.50	442	0.53	624	3.60	610	4.29	1724	1.61	1651	0.19	2088	3.99	2093	4.41
Exchange	442	2.21	417	0.60	673	4.84	646	5.21	1465	2.28	1343	0.47	1979	5.28	1973	5.89
100% read	312	0.01	311	0.01	454	19.71	502	19.90	834	15.6	623	5.58	1654	30.14	1620	27.15

TABLE IX
QUALITY OF SERVICE (IN ms FOR SD AND IN % OF GAIN FOR NASD VS SD) @ TWICE THE RATED ENDURANCE FOR BOTH THE CONSUMER AND THE ENTERPRISE HOST

Workload	Consumer Host								Enterprise Host							
	A-MLC		B-MLC		C-MLC		D-TLC		A-MLC		B-MLC		C-MLC		D-TLC	
	SD	NASD	SD	NASD	SD	NASD	SD	NASD	SD	NASD	SD	NASD	SD	NASD	SD	NASD
MSN	47.07	33.95	32.06	36.32	25.17	22.24	34.58	20.46	53.88	28.32	35.48	34.11	31.62	22.73	45.24	22.34
Financial	14.26	17.85	11.43	22.56	12.36	14.59	13.59	13.00	92.60	37.58	80.65	32.61	71.83	5.25	80.75	26.16
Exchange	7.50	21.35	5.93	14.04	9.37	15.11	8.76	14.00	44.20	22.45	37.83	29.53	39.89	16.46	44.21	23.69
100% read	1.50	23.08	0.77	20.85	2.67	21.47	1.59	29.42	16.20	38.10	15.34	50.80	20.08	43.56	14.40	40.84

TABLE X
NAND FLASH I/O BUS USE (IN % FOR SD AND IN % OF REDUCTION FOR NASD VS SD) @ TWICE THE RATED ENDURANCE FOR BOTH THE CONSUMER AND THE ENTERPRISE HOST

Workload	Consumer Host								Enterprise Host							
	A-MLC		B-MLC		C-MLC		D-TLC		A-MLC		B-MLC		C-MLC		D-TLC	
	SD	NASD	SD	NASD	SD	NASD	SD	NASD	SD	NASD	SD	NASD	SD	NASD	SD	NASD
MSN	99.85	0.01	99.89	0.01	98.28	0.22	98.05	0.33	99.90	0.01	99.93	0.01	98.43	0.19	98.29	0.31
Financial	89.56	0.01	92.26	0.26	74.13	1.53	75.74	0.98	94.90	0.40	96.00	0.13	87.33	1.79	86.52	1.11
Exchange	77.41	0.64	77.17	0.16	62.23	4.03	64.00	3.58	92.63	0.71	95.53	0.26	84.02	5.03	83.29	4.20
100% read	56.53	6.73	45.26	1.82	78.87	12.13	71.14	11.39	94.78	2.43	97.83	0.62	95.78	10.81	95.25	12.64

be scheduled on different dies belonging to the same channel between the two reads required by standard SD.

Tables VII - X show the bandwidth, the average latency, the QoS, and the NAND flash I/O bus use at twice the rated endurance for the 4 tested NAND flash memories and for the two host architectures. NAND flash I/O bus use, sampled with a 1 μs period, is representative of the dynamic power consumption of the whole internal I/O bus. As expected, simulations show that NASD outperforms SD when other commands are scheduled between the two data transfers required by the SD technique, thus temporally separating the data transfer operations and introducing a performance degradation. NASD advantages are highlighted when QoS is concerned since QoS is a metric for worst-case latency conditions rather than an average behavior such as bandwidth and average latency. As it can be observed, the QoS improvements for the MSN workload are in a 20% - 40% range. When looking at the NAND flash I/O bus use, (see Table X), advantages

are materialized only when a 100% random read workload is considered. In fact, when write intensive workloads are devised, the I/O bus transfer time taken by program operations overshadows that of read operations, therefore the reduction in the number of read transfers materialized by NASD is blurred.

IV. CONCLUSIONS

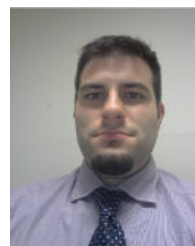
In this paper the potential of a LDPC technique called NAND-assisted soft decision (NASD) is evaluated by comparing its performance with standard LDPC decoding approach. The effectiveness of NASD has been proven through simulations of a 2X MLC, a 1X MLC, a mid-1X MLC, and a mid-1X TLC NAND flash-based SSDs running on a consumer and on an enterprise host system. The results, gathered for synthetic and realistic workloads, show the significant advantages of NASD with respect to the standard approach in particular when the Quality of Service is considered.

REFERENCES

- [1] G. Wong, "SSD market overview," in *Inside Solid State Drives (SSDs)*. Springer-Verlag, 2012, pp. 1–18.
- [2] C. Zambelli, M. Indaco, M. Fabiano, S. Di Carlo, P. Prinetto, P. Olivo, and D. Bertozzi, "A cross-layer approach for more reliability-performance trade-offs in MLC NAND flash memories," in *Design, Automation Test in Europe Conference Exhibition (DATE)*, March 2012, pp. 881–886.
- [3] L. Zuolo, C. Zambelli, R. Micheloni, D. Bertozzi, and P. Olivo, "Analysis of reliability/performance trade-off in Solid State Drives," in *IEEE International Reliability Physics Symposium (IRPS)*, June 2014, pp. 4B.3.1–4B.3.5.
- [4] N. Mielke, T. Marquart, N. Wu, J. Kessenich, H. Belgal, E. Schares, F. Trivedi, E. Goodness, and L. Nevill, "Bit error rate in NAND Flash memories," in *IEEE International Reliability Physics Symposium (IRPS)*, May 2008, pp. 9–19.
- [5] "Solid-State Drive (SSD) Requirements and Endurance Test Method, JEDEC Standard JESD218A," 2011.
- [6] Y. Lee, H. Yoo, I. Yoo, and I. Park, "6.4Gb/s multi-threaded BCH encoder and decoder for multi-channel SSD controllers," in *IEEE International Solid-State Circuits Conference (ISSCC)*, Feb. 2012, pp. 426–428.
- [7] E. Yeo, "An LDPC-enabled flash controller in 40nm CMOS," in *Proc. of Flash Memory Summit*, Aug. 2012.
- [8] X. Hu, "LDPC codes for flash channel," in *Proc. of Flash Memory Summit*, Aug. 2012.
- [9] Erich F. Haratsch, "LDPC Code Concepts and Performance on High-Density Flash Memory," in *Proc. of Flash Memory Summit*, Aug. 2014.
- [10] Tong Zhang, "Using LDPC Codes in SSD — Challenges and Solutions," in *Proc. of Flash Memory Summit*, Aug. 2012.
- [11] K. Zhao, W. Zhao, H. Sun, T. Zhang, X. Zhang, and N. Zheng, "LDPC-in-SSD: making advanced error correction codes work effectively in solid state drives," in *11th USENIX conference on File and Storage Technologies (FAST)*, Feb. 2013.
- [12] R. Micheloni, A. Marelli and R. Ravasio, "Basic coding theory," in *Error Correction Codes for Non-Volatile Memories*, R. Micheloni, A. Marelli, and R. Ravasio, Ed. Springer-Verlag, 2008, pp. 1–33.
- [13] L. Zuolo, C. Zambelli, P. Olivo, R. Micheloni, and A. Marelli, "LDPC Soft Decoding with Reduced Power and Latency in 1X-2X NAND Flash-Based Solid State Drives," in *IEEE International Memory Workshop (IMW)*, May 2015, pp. 1–4.
- [14] "Intel solid-state drive dc s3700 series quality of service." 2013. [Online]. Available: <http://www.intel.com/content/www/us/en/solid-state-drives/ssd-dc-s3700-quality-service-tech-brief.html>
- [15] A. Grossi, L. Zuolo, F. Restuccia, C. Zambelli, and P. Olivo, "Quality-of-Service Implications of Enhanced Program Algorithms for Charge-Trapping NAND in Future Solid-State Drives," *IEEE Trans. on Devices and Materials Reliability*, vol. 15, no. 3, pp. 363–369, 2015.
- [16] Samsung, "Optimized solid-state drives ideal for data center environments," [Online] http://www.samsung.com/semiconductor/global/file/insight/2015/08/PM863_White_Paper-0.pdf, Aug. 2015.
- [17] J. Kim, E. Lee, J. Choi, D. Lee, and S. Noh, "Chip-level raid with flexible stripe size and parity placement for enhanced ssd reliability," *IEEE Transactions on Computers*, 2014, to appear on.
- [18] L. Zuolo, C. Zambelli, R. Micheloni, S. Galfano, M. Indaco, S. Di Carlo, P. Prinetto, P. Olivo, and D. Bertozzi, "SSDExplorer: A virtual platform for fine-grained design space exploration of Solid State Drives," in *Design, Automation and Test in Europe Conference and Exhibition (DATE)*, March 2014, pp. 1–6.
- [19] L. Zuolo, C. Zambelli, R. Micheloni, M. Indaco, S. Di Carlo, P. Prinetto, D. Bertozzi, and P. Olivo, "Ssdexplorer: A virtual platform for performance/reliability-oriented fine-grained design space exploration of solid state drives," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 34, no. 10, pp. 1627–1638, 2015.
- [20] D. Nguyen and F. Roohparvar, "Increased nand flash memory read throughput," Mar. 8 2011, US Patent 7,903,463.
- [21] S. Lee, S. Bae, J. Baek, H. Kim, and S. Kim, "Method of reading data from a non-volatile memory and devices and systems to implement same," Mar. 28 2013, US Patent App. 13/429,326.
- [22] N. Shibata, K. Kanda, T. Hisada, K. Isobe, M. Sato, Y. Shimizu, T. Shimizu, T. Sugimoto, T. Kobayashi, K. Inuzuka, N. Kanagawa, Y. Kajitani, T. Ogawa, J. Nakai, K. Iwasa, M. Kojima, T. Suzuki, Y. Suzuki, S. Sakai, T. Fujimura, Y. Utsunomiya, T. Hashimoto, M. Mikakashi, N. Kobayashi, M. Inagaki, Y. Matsumoto, S. Inoue, Y. Suzuki, D. He, Y. Honda, J. Musha, M. Nakagawa, M. Honma, N. Abiko, M. Koyanagi, M. Yoshihara, K. Ino, M. Noguchi, T. Kamei, Y. Kato, S. Zaitou, H. Nasu, T. Arika, H. Chibvongodze, M. Watanabe, H. Ding, N. Ookuma, R. Yamashita, G. Liang, G. Hemink, F. Moogat, C. Trinh, M. Higashitani, T. Pham, and K. Kanazawa, "A 19nm 112.8mm² 64Gb multi-level flash memory with 400Mb/s/pin 1.8V Toggle Mode interface," in *IEEE International Solid-State Circuits Conference (ISSCC)*, Feb. 2012, pp. 422–424.
- [23] D. Lee, I. J. Chang, S.-Y. Yoon, J. Jang, D.-S. Jang, W.-G. Hahn, J.-Y. Park, D.-G. Kim, C. Yoon, B.-S. Lim, B.-J. Min, S.-W. Yun, J.-S. Lee, I.-H. Park, K.-R. Kim, J.-Y. Yun, Y. Kim, Y.-S. Cho, K.-M. Kang, S.-H. Joo, J.-Y. Chun, J.-N. Im, S. Kwon, S. Ham, A. Park, J.-D. Yu, N.-H. Lee, T.-S. Lee, M. Kim, H. Kim, K.-W. Song, B.-G. Jeon, K. Choi, J.-M. Han, K. H. Kyung, Y.-H. Lim, and Y.-H. Jun, "A 64Gb 533Mb/s DDR interface MLC NAND Flash in sub-20nm technology," in *IEEE International Solid-State Circuits Conference (ISSCC)*, Feb. 2012, pp. 430–432.
- [24] PCI-SIG Ass., "PCI Express Base 3.0 Specification," [Online]. Available: <http://www.pcisig.com/specifications/pciexpress/base3/>, 2013.
- [25] "Nvm express 1.1 specification," 2013. [Online]. Available: http://nvmeexpress.org/wp-content/uploads/2013/05/NVM_Express_1_1.pdf
- [26] "Hp z640 workstation," 2015. [Online]. Available: <http://www8.hp.com/h20195/v2/GetDocument.aspx?docname=c04434085>
- [27] "Flexible I/O tester," 2015. [Online]. Available: <http://freecode.com/projects/fio>
- [28] X.-Y. Hu, E. Eleftheriou, R. Haas, I. Iliadis, and R. Pletka, "Write amplification analysis in flash-based solid state drives," in *Proceedings of SYSTOR 2009: The Israeli Experimental Systems Conference*, 2009, pp. 10:1–10:9.
- [29] S. H. Shin, D. K. Shim, J. Y. Jeong, O. S. Kwon, S. Y. Yoon, M. H. Choi, T. Y. Kim, H. W. Park, H. J. Yoon, Y. S. Song, Y. H. Choi, S. W. Shim, Y. L. Ahn, K. T. Park, J. M. Han, K. H. Kyung, and Y. H. Jun, "A new 3-bit programming algorithm using SLC-to-TLC migration for 8MB/s high performance TLC NAND flash memory," in *Symp. on VLSI Circuits*, Jun. 2012, pp. 132–133.



Lorenzo Zuolo received the Laurea Magistrale degree (M.Sc.) and Ph.D. in electronic engineering from the University of Ferrara, Ferrara, Italy, in 2012 and 2016, respectively. Currently, he holds a Research Assistant (post-doctoral) position in the Dipartimento di Ingegneria of the same institution. His main research interests are focused on architectural/physical simulation of Solid State Drives (SSD) and emerging non-volatile memories.



Cristian Zambelli received the M.Sc. and Ph.D. (Hons.) degrees in electronic engineering from the University of Ferrara, Ferrara, Italy, in 2008 and 2012, respectively. He has held a Research Assistant (post-doctoral) position with the Department of Engineering, University of Ferrara, since 2012, where he is currently an Assistant Professor. His current research interests include the characterization, physics, and modeling of nonvolatile memories reliability and solid state drives reliability.



Alessia Marelli is Senior Design Engineer Microsemi. Before Microsemi she joined Integrated Device Technology (IDT) in 2009 as senior digital designer, where she took care of ECC applied to SSD. In 2007, she joined Qimonda as senior digital designer. From 2003 to 2007 she joined STMicroelectronics, Agrate B., Italy where she was involved in digital design of Multilevel NAND Memories, especially redundancy, ECC and algorithms. She is co-author of Memories in Wireless Systems (Springer, 2008), Error Correction Codes for Non-Volatile Memories (Springer, 2008), Inside NAND Flash Memories (Springer, 2010) and Inside Solid State Drives (Springer, 2013).



Rino Micheloni is Fellow at Microsemi. Before Microsemi, he was Lead Flash Technologist at IDT (Integrated Device Technology), Senior Principal for Flash and Director of Qimonda's design center in Italy, developing 36 nm and 48 nm NAND memories. From 2001 to 2006 he managed the Napoli design center of STMicroelectronics focusing on the development of 90 nm and 60 nm MLC NAND Flash. Before that, he led the development of MLC NOR Flash.



Piero Olivo received the Ph.D. in electronic engineering from the University of Bologna, Bologna, Italy, in 1987. He has been a Full Professor of Electronics with the University of Ferrara, Ferrara, Italy, since 1994. His research interests include the physics, the reliability and the experimental characterization of innovative non-volatile memory cells and architectures.