# FLASH STORAGE IN A LINUX-BASED RESIDENTIAL GATEWAYS

Zachi Friedman

M-Systems, Inc.

8371 Central Ave., Newark CA, 94560

zachi.friedman@m-sys.com

**Abstract**

The availability of new in-home interconnection technologies, combined with the explosion of non PC-based devices, is driving the demand for a single device to connect in-home appliances to the public Internet. The Residential Gateway (RG), as its name implies, is a central entry and control point at the home for current voice, video and data services, as well as the cornerstone for future services. It is the true enabler of the Age of Information.

Embedded Linux distributions play a leading role in RGs. Memory is among the four major RG building blocks. Flash memory is used in many RG designs to safeguard the gateway's OS, and store both user and network data. Because of this, it is essential to understand the special considerations that Linux designers face when selecting and working with flash memory.

This paper begins with an overview of the RG market and its major components. It then discusses the popular flash storage options available today, describing their different characteristics and the markets to which they bring the most benefits. It addresses the concerns of ~40% of the respondents to a recent poll who said: "My main concern about using Linux in embedded applications is insufficient driver support from chip vendors," as well as special considerations among designers when choosing the appropriate flash storage for a Linux-based RG. This includes how to choose a flash file system, and how to avoid data/code corruption and premature flash block expiration. It presents a case study that provides a real-life example of a Linux-based RG, demonstrating how to future-proof a successful, Linux gateway design as perceived by M-Systems based on its experience with leading Original Equipment Manufacturers (OEMs) and Original Design Manufacturers (ODMs).

## Introduction

The Residential Gateway (RG), or home gateway, sits between the Wide Area Network (WAN) and the Small Office Home Office (SOHO) Local Area Network (LAN), serving as the core of the home network. It enables bi-directional communication and data transfer among networked appliances in the home and across the Internet, and serves as an access platform for service providers to remotely deploy services to the home. Embedded Linux plays a major role in RG architectures, side-by-side with traditional OSs such as Wind River's VxWorks. Recent RG designs are more feature-rich and, consequently, require higher flash storage capacity to store the OS image, applications, and system and user data.

A number of factors are driving the development of the RG market:

- Increased availability of new home networking technologies
- Expansion of service offerings by network operators
- Increased use of the Internet
- Increased demand for non PC-based appliances
- International standards
- Increased availability of broadband
- Availability of new entertainment options, including digital television and streaming multimedia
- Need for increased security and protection

## The Building Blocks of a Residential Gateway

A typical RG consists of:

- A Central Processing Unit (CPU)
- Local memory
- A chipset and digital modem
- Software

### CPU

The CPU is the central component of the RG, responsible for almost every task that it performs. The CPU can affect many system characteristics, including: the OS, applications, power consumption, system stability, Bill Of Materials (BOM). Some of today's CPUs are actually Systems-on-a-Chip (SoC), all-in-one products that integrate the networking components and the CPU into one cost-effective chip. Such devices or derivatives are widely available from Broadcom, Conexant, Texas Instruments and National Semiconductor.

### Local Memory

A gateway requires local memory to store and manipulate instructions issued by the subscriber or the operator. The gateway uses two main categories of local memory, volatile and non-volatile. RAM is volatile; i.e., it loses its contents when power is turned off or fails. RAM is mainly used to store and access application code. The more RAM available, the faster the gateway responds. Flash memory and hard disks are non-volatile memory. They safeguard the gateway's OS and customizable features, as well as user data, network data and many other elements. Each of these types of non-volatile memory can benefit an RG in different ways, as described in the next section.

### Chipset and Digital Modem

The chipset, located on the home networking (LAN) side of the RG, provides the interface to the particular technology running on the network. The modem, usually DOCSIS compliant, is located on the broadband (WAN) side.

### Software

The RG software consists of the OS and the applications running on top of it. The software should enable the smooth inter-operation of information appliances and services within the home. This requires, first and foremost, OS stability. In addition, the OS should be compact, provide a responsive kernel, and be modular to enable frequent and easy upgrades to future-proof the RG.

Among the applications running on top of the OS are: Electronic Programming Guides (EPGs), real-time decoders, web browsers, games, messaging software and VoIP software.

## Non-Volatile Memory in a Residential Gateway

Non-volatile memory can be subdivided into two types: solid-state flash memory and mechanical hard disks.

Hard disks are available in higher capacities than flash memory and at a lower cost per Megabyte. But the mechanical nature of hard disks both shortens their life span and can cause reliability problems in home environments. This can be worsened by the typically constrained environment of consumer electronics device enclosures. In addition, the Total Cost of Ownership (TCO) of maintaining hard disks in the living room is often overlooked. High return rates due to disk failure – some set-top box manufacturers report an annual failure rate of between 2 to 8% – can financially damage the business model of operators delivering home services, as well as the manufacturers' reputation and brand name.

This explains why many hardware engineers are incorporating both flash memory and hard disks into their design to meet the storage needs of next-generation RGs. This type of hybrid memory is an ideal model. It can provide both the high capacity and cost-effectiveness of a hard disk and the reliability of flash memory, enabling advanced capabilities such as PVR and MP3 players while providing reliable operation over time. In the event of disk failure, the mass storage recording and viewing would be disabled, but the gateway would continue to perform its other functions, such as broadband Internet, web browsing, instant messaging and VoIP.

**M-Systems' DiskOnChip**

## Flash Memory

### NOR and NAND Flash Comparison

NOR and NAND technologies dominate today's solid-state, non-volatile flash memory market. Both technologies have unique features and are aimed at fulfilling different market needs. NOR technology has evolved from the historic ROMs, PROMs, and EPROMs, basically replaced them in capacities over 8Mbits. It is typically used for code storage and execution purposes only, and is thus primarily used in simple digital devices such as low-end, voice-centric cell phones or simple cable modems without any real gateway connections.

NAND flash, a newer technology, was introduced to meet the demand for much higher capacity data storage at much lower prices than NOR. It is often used for both code and data storage in devices such as set-top boxes, MP3 players, digital cameras, high-end smartphones, PDAs and gateways.

It is crucial to understand that raw NAND cannot replace NOR unless coupled with a boot ROM and a controller to overcome raw NAND's inherent limitations, and/or special software to manage and correct data errors. This is because NAND was originally intended for use with media files, such as JPEGs and MP3s, where an occasionally flipped bit does not compromise the application. But even this is changing today as compression becomes key.

From a performance perspective, NOR is optimized for reading data but significantly lags behind NAND performance when writing and erasing data. This often disqualifies the use of NOR in RGs. Typically, NAND outperforms NOR in such operations by orders of magnitude.

By examining the physical architecture of both technologies, NAND offers higher densities with more capacity on a given die size, thus making its cost structure far more attractive (anywhere between 2 to 4X!). This, in combination with a

simpler production process, enables manufacturers to build NAND products with a capacity range of 64Mbits to 1Gbit. In addition, since NAND flash need not be perfect (usually up to 2% of the blocks may be bad), its production yield is much higher, again resulting in a significant cost reduction.

## Flash Data Reliability and Management

Unlike NOR flash which is a perfect media, NAND flash has inherent reliability problems. The three most important obstacles to using NAND flash for reliable storage are:

- Bad Block Mapping – Since up to 2% of flash blocks may potentially be bad, designers must implement a mechanism to identify these blocks. Once identified, bad block information must be permanently stored to prevent these blocks from being accessed when reading from the flash or writing to it.
- Error Detection and Error Correction (EDC and ECC) – Flash manufacturers' specifications state that designers should expect bit errors (bit flipping) of up to 1 bit per page. Therefore, flash designers must employ a mechanism to identify and correct at least this number of bit errors per page.
- Wear-Leveling – Both NAND and NOR flash have a limited number of erase cycles per block, after which degradation of reliability and performance can be expected for operations in that block. Designers must use wear-leveling algorithms to systemically distribute data equally among all blocks rather than among the same blocks, thereby maximizing reliability and extending flash life span to support designs in the field.

Flash designers use these strategies to overcome flash shortcomings and to manage it:

- An external hardware controller – Interacts with the flash array to both overcome raw flash reliability problems and to provide the functionality of a mechanical hard drive on a solid-state silicon chip. The performance level for this type of NAND management system is relatively high; however, the cost structure associated with its implementation is obviously higher than stand-alone raw NAND, and its overall endurance and reliability is questionable at best.
- Management software – Typically, a NAND flash management software system uses intricate code that runs on the host CPU. Although the low cost of implementation is enticing (not taking into account software engineering efforts), severe performance and reliability penalties plague this software-only solution, making it a problematic design choice. Every change in OS, flash capacity or flash type carries with it hidden costs, due in part to the need for new drivers and algorithms, making this choice even less attractive. Also, software-only solutions do not work with newer flash technologies such as Multi-Level Cell (MLC) NAND, the most cost-effective flash to be introduced in 2003 by Toshiba.
- A mixed balance of hardware and software – A hybrid solution, such as M-Systems' DiskOnChip, uses a balanced combination of a thin controller and a software driver. The controller is embedded into the same silicon die as the flash array itself, and performs computational-intensive tasks on-the-fly, with minimum increase to the cost of raw NAND flash. The software is actually a block device driver, which exports sector read and write operations to the OS. A hybrid solution provides high-reliability and performance levels, in combination with an attractive price. It is also able to address new technologies, such as MLC NAND, for an even better cost structure than raw traditional Bi-Level Cell (BLC) NAND.

The separate and incompatible paths that Toshiba and Samsung plan to take over the next few years in their efforts to reduce of flash cost will affect these strategies. Toshiba's MLC NAND flash will drive the cost of NAND down even further. Samsung's BLC NAND flash will decrease the silicon size by shrinking the geometry of the flash production process. A direct hybrid solution, such as M-Systems DiskOnChip, will enable designers to use the same driver to support both of these technological paths. This further increases the value of a hybrid solution.

### NOR and NAND Feature Summary

The table below compares the basic features of NOR, NAND and NAND-based DiskOnChip.

| Feature | NOR | NAND | NAND-Based DiskOnChip |
|---|---|---|---|
| Available capacities | 8-64Mb in SLC<br>64-256Mb in MLC | 64-1024Mb in BLC<br>256-1024Mb in MLC (Q1, 03) | 128-512Mb in monolithic BLC<br>512Mb in monolithic MLC (Q1, 03) |
| No. of Write/Erase cycles | 10,000 – 100,000 | 100,000 – 300,000 | 1,000,000 statistically (includes wear-leveling, bad block management and EDC/ECC) |
| Requires Bad Block Management | No – perfect media | Yes – Implemented with software and/or external controller | Yes – Implemented with built-in hardware and software combined solution |
| Read/Write sustained speeds | Read – >3 MB/s<br>Write – 100KB/s<br>(16-bit width) | Read – 1.5 MB/s<br>Write – 0.8 MB/s<br>8-bit width only! | 8-bit R/W – 1.5/0.8 MB/s<br>16-bit R/W – 3.1/1.7 MB/s |
| Execute in Place (XIP) | Yes | No | 1KB programmable boot block with XIP capability |
| Requires additional boot ROM | No | Yes | No |
| Hardware integration | Easy interface | Multiplexed – not as easy to integrate | Easy SRAM-like interface |
| EDC | Optional – in software | Mandatory – software or external controller | Built-in hardware EDC and software ECC |
| Other features:<br>Unique ID<br>OTP area<br>Hardware protection | <br>No<br>No<br>Partial (Intel StrataFlash) | <br>No<br>No<br>No | <br>Yes<br>Yes<br>Yes |
| Cost structure | ~2-4X | 1X | ~1.1-1.2X<br>~0.7X for MLC |

**TABLE 1***: Flash Feature Comparison*

## Choosing the Right Non-Volatile Memory for a Residential Gateway

The appropriate storage capacity for an RG project ranges from 8Mbit to 512Mbit for flash, to tens of Gigabytes for hard disk drive storage.

The minimalist, procurement-driven 8-32Mbit approach carries little added marketing value, is not future-proof, and provides limited functionality. A true RG must provide room to deliver a myriad of services (current and future) to subscribers. The minimal flash capacity required is 128Mbits for today's designs and between 256Mbits to 512Mbits for future designs.

When choosing flash technology, NOR is suitable if the design requires 32Mb or less. However, if 64Mb or more is needed, NAND is the most cost-effective solution.

## Software in a Residential Gateway

### Operating System

Although there are several OSs of choice for a new RG design, Linux is becoming more and more popular for several reasons:

- No licensing fees or royalties – unlike Windows CE.NET or VxWorks
- Open source – Linux provides designers with a wide range of customization capabilities
- An extensive code-base – for device drivers, kernel tweaks and patches, a vast selection of file systems, and many freely available applications (specifically for RGs: EPG software, PVR software, browsers, e-mail clients)

The following quotation from Linux Devices supports this analysis:

"The Linux operating system is well suited for use in the rapidly growing embedded computing market. It's technologically advanced kernel, open source development model, free availability and royalty free distribution make it an ideal choice for future designs. The large developer environment and fast pace of contributions ensure that Linux will meet the requirements of emerging embedded and mobile applications for some time to come." [Greg Haerr, LinuxDevices.com (Oct. 5, 2001) – http://www.linuxdevices.com/articles/AT7695438395.html.]

## Device Driver

If Linux is the OS of choice and M-Systems' DiskOnChip is the local flash memory used, several device driver options are available. The device driver resides between the file system and the hardware, causing the file system to "see" the flash memory as a hard disk-like block device.

There are three major device driver options for a device driver: proprietary, MTD driver, M-Systems' TrueFFS® driver. The following table compares these options.

| **Feature** | **Proprietary** | **MTD Driver** | **TrueFFS** |
|---|---|---|---|
| R&D effort (time) | Months | Hours-days | Hours-days |
| Flash supported | Depending on needs | NOR, NAND, most DiskOnChip products | DiskOnChip only |
| File system supported | Depending on needs | Block device<br>JFS | Every file system that works with a block device driver (ext2, ext3, etc.) including journaling file systems (JFS, XFS, etc.) |
| Maintenance effort | Update needed with every flash change (either capacity or manufacturer) | Update needed by MTD group | None – each new DiskOnChip is supported with a driver update |
| MLC NAND technology support | Complete rewrite | Major updates/ rewrites needed | Available with first MLC DiskOnChip (Q1, 03) |
| Error Detection Code (EDC) | Depending on needs | Pure software, performance degradation | Hardware-based and on-the-fly, no performance penalty |
| Reliability | Probably would become stable within a few years | Quite stable | Stable –algorithms constantly improved over the last 13 years |
| Bad Block Management (mandatory for NAND-based flash) | Coding needed | Done | Done |
| Wear-Leveling | Coding needed | Working solution | Working solution implementing both dynamic and static wear-leveling |
| Support | None | MTD group must support through mailing list | M-Systems' free support through phone, e-mail and on-site |
| GPL licensing | Depending on needs | Yes | Yes |
| Source code availability | Yes | Yes | Yes – under a license agreement |

**TABLE 2***: Device Driver Comparison*

## A Future-Proofed, Linux-Based Residential Gateway

The figure below shows a real-life, typical gateway installation, a Broadcom wireless DOCSIS cable modem gateway based on BCM 3360, or BCM 3348 SoCs with DiskOnChip Millennium Plus 128Mbit (16MByte) flash.

Because broadband technology is relatively new and continuously evolving (DOCSIS 2.0, HPNA 2.0, Home Plug, Cable Home, xDSL), the functionality of an RG must be upgradeable to insure design relevancy over the next 3 to 7 years. This implies building a modular, future-proof hardware and software design. Currently, quite a few of Broadcom's new reference designs use Linux on MIPS-based and PowerPC processors. These reference designs use M-Systems' DiskOnChip Millennium Plus 128Mbit and/or 256Mbit devices as their preferred local flash storage solution. Broadcom has chosen to use M-Systems' TrueFFS driver, which exports a standard block device API into the file system and OS. This enables ODMs and OEMs to work with whichever file system they choose.

Other CPU architectures may play an important role in the RG, namely x86-based CPUs, or SoCs from National Semiconductor, VIA and SiS. These gateways may offer more PC-like functionality. Linux is also likely to be a popular choice for x86-based designs.

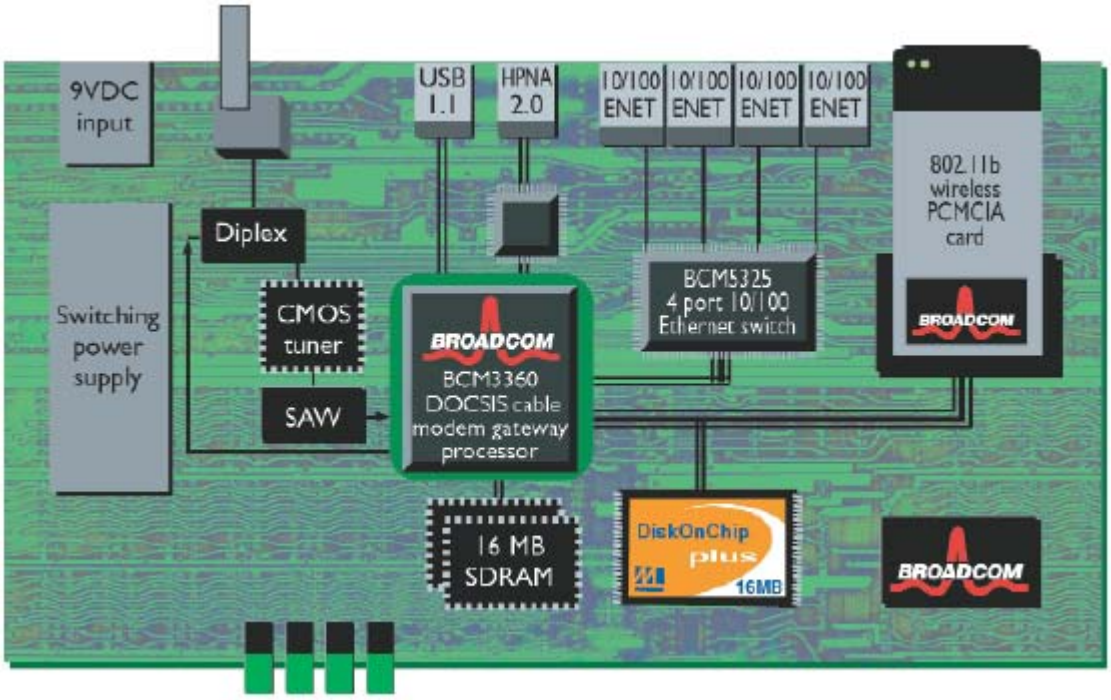


**FIGURE** 1*: Broadcom wireless DOCSIS cable modem gateway*

## Flash Trends

Two main trends in the flash industry will affect the role flash memory plays in future Linux-based RG installations:

- Increased capacity by adding more data per cell, as with MLC technology. MLC NOR flash, used in Intel's StrataFlash, was the first to move in this direction. NAND products based on Toshiba's MLC technology will be available in new M-Systems' DiskOnChip products in early 2003.
- Increased capacity by shrinking the flash process (smaller geometry). Currently the flash industry is moving toward 0.12-0.13µ production processes.

The momentum that these trends gain within the embedded community at large and within the Linux community, specifically, will depend in large part on how cost-effectively, reliably and easily they offer future-proof solutions to the growing storage needs of RG service providers.