SPAY020—November 2004
White Paper

Delivering Cost-Effective Voice-over-Cable Modem Without Compromising Future Extensibility

Frank Fruth, Piyush Patel and Oded Melamed

Texas Instruments

Introduction

As cable-based, voice-over-IP trials prove successful, it has become clear that deploying voice-over-cable modem (VoCM) services is the next major revenue source for Multiple System Operators (MSO). One of the keys to a profitable rollout of VoCM is robustly integrating all of the functions a customer needs today onto a single CPE hardware platform. The primary engineering challenge is balancing the cost of the CPE against future expandability.

The CPE platform is often referred to as an embedded multimedia terminal adapter (EMTA). An EMTA combines delivery of high-speed data with voice-over-IP services by connecting legacy telephones and terminal equipment (like a fax machine) to a cable operator's advanced IP communications network. A typical EMTA has one or two RJ-11 telephone ports for either a direct phone connection or to provide voice connectivity through all wall jacks in the home or office, as well as USB and Ethernet ports for data connectivity.

The minimum range of telephony services that must be supported by an EMTA is defined by the PSTN devices to which EMTAs connect. These services include basic phone functions such as call waiting, three-way calling, voice mail and also fax support. In addition, MSOs must provide enhanced functionality to differentiate their products. The EMTA is the MSO's revenue gateway, and any limitations of the device could limit profit. Therefore, an EMTA that provides only marginal voice quality or has difficulty interoperating with other telephony devices will not have long-term market viability. Certainly there is pressure to provide the lowest cost EMTA today to accelerate market penetration, but this needs to be tempered with user expectations and an agility to keep equipment compatible with next-generation devices and services.

Understanding the Basics of an EMTA

Today's standard EMTA provides 2+2 voice service (two lines of telephone service with call waiting/conferencing capabilities on each line) with 8ms line echo cancellation. However, the phone network carries more than just voice traffic, such as fax and data modem traffic, and EMTAs must support these "legacy" services as well.

EMTAs can support fax services through the T.38 fax relay specification from the International Telecommunications Union (ITU). Using T.38, an EMTA can detect facsimile calls and convert analog facsimile signals to digital data suitable for transmission over a packet network. The transmission of a digital representation of facsimile signals not only reduces the overall network bandwidth requirements but also allows for improved transmission reliability by employing mechanisms specific to fax for compensating packet network impairments.

In addition, there are a number of home applications that employ voice-band data modems like security alarms and TIVO. Packet loss within the IP network can reduce the overall performance and reliability of these services. Consumers will resist purchasing VoCM services if they compromise home security or the entertainment experience. In order to address the reliable handling of home voice-band data modem applications, the ITU is in the process of defining a V.VBD (Voice Band Data) specification providing the means for using a G.711 PCM codec bolstered with forward error correction (FEC) and redundancy the mechanisms (RFC2198 and RFC2733) to increase network robustness to an acceptable level. EMTA will be expected to implement techniques defined by the V.VBD standard to ensure PSTN-like reliability of home voice-band data modem applications.

All this effort merely achieves parity with incumbent voice service providers. The challenge for designers is providing enhanced features beyond the minimum functional requirements in order to differentiate their products while managing cost. Current T.38 implementations, for example, support only traditional fax modem protocols with transmission rates up to 14400 bps. As the cost of higher quality V.34 fax machines supporting transmission rates up to 33600 bps continues to drop, market analysts predict that they will comprise the overwhelming majority of devices shipped in 2006. The ITU has defined a new version of T.38 to incorporate support for V.34 facsimile protocols. EMTAs not supporting this new version of T.38 will force V.34 devices to operate at a reduced level of functionality. EMTAs will need to support V.34 fax relay to support these devices if they don't want to lose market share to those EMTAs that can.

It is important to remember that the primary reason EMTAs have to provide more than the minimum set of telephony services is to provide differentiation over the PSTN. Offering an improved user experience will provide the required motivation for customers to switch off the PSTN and switch on VoIP. For example, one limitation in the PSTN is its 8-KHz sample rate which limits bandwidth of voice signals. VoIP in general is not bound by this limitation and therefore has the ability to provide enhanced call quality through the use of wideband codecs.

Voice and the EMTA

In the telephony world, each time a voice packet is translated from one codec format to a second format (a process called transcoding), voice quality is degraded. A call between a cell phone and EMTA undergoes at minimum two transitions (see Figure 1a) when packets are, for example, converted from/to GSM and to/from G.729. Compression is lossy, and the more transitions there are – especially as calls pass through different service providers and networks – the lower the overall quality.

Standard codecs pass through a rigorous certification process, guaranteeing that loss of quality is controlled and within acceptable limits. However, when you transcode multiple times, losses compound to a point where users can begin to perceive them. Consider the drop in quality in a cellular-to-EMTA connection with minimum transcoding (see Table 1). GSM-AMR baseline has an average PESQ of 3.997, whereas GSM-AMR to G.729AB has a score of 3.444. The difference in quality is 0.553, which is unacceptable since a drop of only 0.1 is noticeable by most users. With more transcoding, the loss becomes even more pronounced.

Adoption rates of EMTA services will be negatively impacted if users perceive a loss of quality compared to existing PSTN services. As an alternative, to avoid transcoding losses, EMTAs have an opportunity to eliminate the use of codecs in tandem (see Figure 1b). Instead of compressing all voice traffic using a limited variety of cable-centric codecs that are transcoded at network media gateways, EMTAs can increase call quality by supporting the native codecs of the devices with which they interact. For example, when connected with a mobile device, an EMTA could shift from a G.729 codec to the GSM-AMR codec used natively by the cell phone. In this way, transcoding is eliminated as are all associated losses in quality, and EMTAs can enjoy a quality advantage over legacy telephone service.

Tandem free operation is an essential characteristic of convergence devices and will be a major differentiator between MSOs and incumbent carriers. By accommodating each end device a user wishes to connect to, the EMTA guarantees the highest call quality possible; by negotiating directly with the cell phone rather than the media gateways between them, the EMTA is able to select the best codec to use.

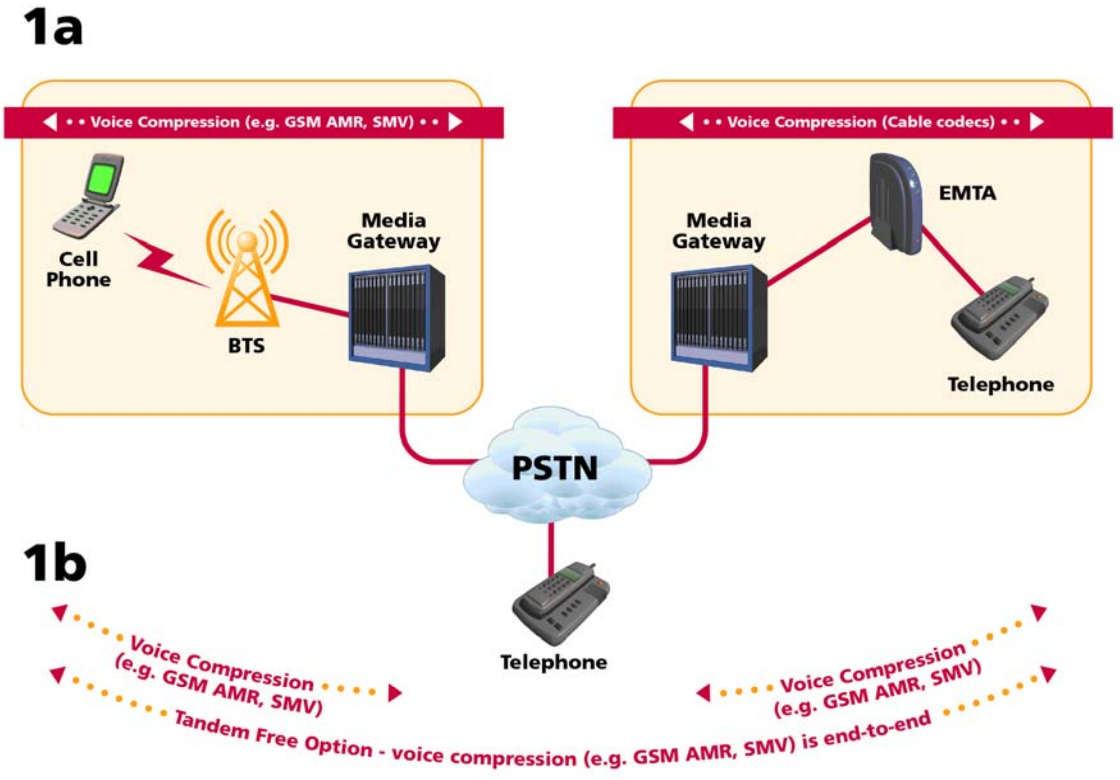

Figure 1: Tandem Free Operation
A cellular-to-EMTA call undergoes a minimum of two transitions (a) when voice encoded at the EMTA using a cable voice codec is converted to a cellular codec like GSM. Such transcoding between codecs introduces perceivable loss of voice quality. By employing tandem free operation (b) where an EMTA uses the native codec of the device it is connected with, EMTAs avoid multiple compression and decompression stages and maintain maximum call quality.

Loss of Voice Quality Due to Transcoding

	Average PESQ
SMV (baseline)	3.667
GSM-AMR (baseline)	3.997
G.729E (baseline)	3.613
G.729AB (baseline)	3.994
SMV to G.729AB	3.230
G.729AB to SMV	3.228
G.729AB to GSM AMR	3.444

GSM AMR to G.729AB	3.397
SMV to G.729E	3.461
G.729E to SMV	3.453
GSM AMR to G.729E	3.724
G.729E to GSM AMR	3.700

Table 1: Transcoding voice through several codecs compounds the reduction of voice quality from lossy compression. Compare the quality of an EMTA-to-cellular connection between an EMTA using G.729 AB (3.444 PESQ) rather than native GSM (3.997). Note: Most users can perceive a drop in average PESQ as low as 0.1.

The Architecture of an EMTA

Given that an EMTA can determine the best codec to use for a particular call, what remains is to enable the EMTA to support a rich set of codecs beyond the basic minimum required. The right architecture for a given application must optimize processing capacity, ease of programming, memory footprint, board size, and other characteristics that directly affect overall system cost.

An EMTA architecture is comprised of four main subsystems: DOCSIS® cable modem, application processor, VoCM processor and an analog subsystem. Typically, the cable modem and application processor can be implemented on the same processor integrated with the appropriate peripherals. VoCM processing should be implemented using a programmable device so that the EMTA can be upgraded as new standards, codecs and services become available.

For VoCM processing, DSPs offer superior performance over general purpose processors (GPP) like MPUs and MCUs. Execution of the critical loop in a voice codec uses significantly fewer cycles when implemented on a DSP because of inherent efficiencies in DSP architectures, including highly parallelized pipelines, an optimized instruction set and specialized compute units.

One of the most prominent characteristics of DSPs is that they are designed to provide single-cycle multiplication and accumulate (MAC), a function used extensively in signal processing algorithms. GPPs that attempt to shoe-horn in MAC functionality may provide a single-cycle multiplication instruction but don't have the native architecture to exploit the instruction to its full capabilities. For example, most GPPs are based on a von Neumann architecture. The von Neumann architecture provides a single memory access per instruction cycle. A MAC function, however, requires several memory accesses; for a FIR filter for example, the processor must fetch the MAC instruction, read the next coefficient, get the next data sample from the delay line, and write back the current data sample to the delay line. While the GPP MAC instruction alone may take a single clock cycle, it must be supported with several memory accesses.

DSPs are based on what is known as a Harvard architecture, which supports multiple memory spaces, enabling all multiplication, accumulating and memory accesses to take place in a single cycle. The update of the delay line on DSPs has almost no overhead because of hardware support for circular addressing, whereas GPPs must expend cycles to directly manage the delay line. Zero overhead loop (discussed below) is also suitable for reducing the FIR processing cycles. As a result, even though both GPPs and DSPs may support single instruction multiplication, a GPP will require several cycles for each tap in our filter example to a single cycle on a DSP. Relatively newer DSP architectures like TI's TMS320C55x™ processor even support two multiply-accumulates in one cycle with no penalty due to the extra memory access required to support this enhancement, which gives a boost to the MAC-intense code that dominates the processing requirements.

Another architectural difference involves loop management. Since most signal processing takes place in a short critical loop, DSPs have been architected with special hardware to provide zero-overhead looping, meaning that managing the critical loop consumes no cycles. On a GPP, the programmer must test the loop counter, conditionally jump to the beginning of the loop and adjust the loop counter. This overhead occurs for every iteration of the loop.

Since DSPs like TI's C55x™ processor have a complex instruction set with dense encoding, the code size required to accomplish the same task is about 1.5 to two times less than on a GPP. This results in smaller instruction cache size and less penalty when accessing slow external memories, which in turn gives higher performance at reduced cost.

In GPPs, architectures are not optimized in the same way as a DSP, resulting in two to three times more CPU cycles to accomplish similar signal processing tasks. Put another way, it takes a GPP running at 300 MHz to execute the equivalent processing on processors tuned for digital signal processing running at 100 MHz.

Note that the limiting factor in implementing extended features is available headroom. Processor cost does not increase linearly with clock speed, so a GPP with more headroom will increase in cost faster than an equivalent DSP. Additionally, if a processor is already at its maximum clock speed, it has no headroom available.

In some GPPs, it is possible to extend the instruction set to improve signal processing performance. Such extensions do not solve the fundamental problem of multiple access per cycle. They require design and verification effort and are usually not optimal in power, area or performance because they are not integral to the processor pipeline. Even with such extensions, it may be possible to meet only a basic set of requirements. Some recent announcements made by GPP vendors promise to deliver signal processing performance by having SIMD extensions integral to the processor pipeline. Although they improve memory access bandwidth, these extensions still do not completely solve the multiple access issues. In addition, these extensions are far less mature with respect to DSP software support and silicon compared to traditional DSPs and thus have a disadvantage in terms of time-to-market. Further, improvements are effective only if computations allow memory loads to have high reuse within the constraints of available registers. For the FIR example, it usually takes significant loop unrolling and careful instruction scheduling to approach the gains that are naturally achievable with compact code on traditional DSPs. The extra unrolling results in a higher program code size which may also require higher cache sizes, in line with the overall code size growth mentioned earlier.

Quality is Key

Improving quality is a never-ending challenge for designers, and there is always ample room for improvement. For example, delays in the PSTN are small, so the need for echo cancellation there is not great. However, with the larger delays associated with a packet network, echo cancellation plays a much more important role in overall voice quality. Echo cancellation, however, comes at the cost of consuming signal processing resources.

For years, PCM has been a key codec for telephony applications because it provides high voice quality, ensures interoperability and requires little processing overhead. The primary problem with PCM is its cost in terms of bandwidth. G.711 PCM requires 64 kbps per channel, compared to a low bit rate codec (LBRC) like G.729, which provides sufficient voice quality at a bit rate as low as 8 kbps. In economic terms, one can support eight calls of G.729 for every call of G.711.

While conserving network bandwidth and significantly lowering network operational costs, LBRCs require more processing resources on the EMTA. Typical EMTA configurations today allocate an LBRC for main line use and use PCM for conferencing/call waiting services. DSPs, because of their signal processing nature, provide a cost-effective means for enabling an EMTA to support 2+2 operation using LBRCs for both the main line voice service as well as for the conferencing/call waiting services, enabling maximum bandwidth efficiency.

Reliability, Dependability

For an EMTA to become the sole piece of telephony equipment at the customer premise, it must provide the same guaranteed dial-up service during loss of power to the home that PSTN-based phones provide. Without power, a voice-enabled cable modem will not be able to complete calls.

Guaranteed service can be maintained through the addition of a battery pack and corresponding charging/monitoring circuitry. Providing battery backup is a key requirement if MSOs want to offer a true Local Exchange Carrier (LEC) service. Battery backup guarantees service availability and also allows the EMTA to remain operational during momentary power events such as brownouts or power spikes. Creating a battery-backed architecture, however, requires a mature power design strategy. For example, Lithium Ion battery packs require charge-controlling circuitry because their voltage is not constant over the useful range of a battery's life. They also lose capacity after a large number of charge cycles. TI provides the circuitry, firmware and design guidance required to address and solve these issues efficiently.

While power events are not typical, they must be handled eloquently. Designing with low power consumption in mind will improve the reliability of an EMTA during extended power events. For signal processing functions, DSPs have significant power efficiency advantages over GPPs because GPPs must be clocked at a higher frequency to achieve the equivalent processing on a DSP. Clocking at a higher frequency results in a corresponding increase in power consumption and lower overall battery life.

The Converged Home Network

Integrating the wireless LAN (WLAN) within the EMTA provides a foundation for exciting new home networking applications. Broadband-enabled phones will support high-quality audio, live video streaming, full-speed web access, automatic and instantaneous synchronization with address books and e-mail services, and real-time Internet gaming.

An EMTA's ability to support voice-over-WLAN also enables convergence of the cable and cellular networks. Users will use the traditional cellular network when they are mobile and access robust, higher quality voice service over the cable broadband IP network when they are home or at the office, all using the same dual-mode cellular/ Wi-Fi® handset.

The value proposition from such convergence promises to have far-ranging impact. Cellular-over-cable modem will offload residential tower demand. Cellular/landline convergence will also enable subscribers to collapse their phone service to a single number that can reach them at the office, home, or on the road, with automatic find-me/follow-me, a single voice mailbox and dialing directory, and other enhanced services. Ultimately, these features mean less churn and higher revenues for MSOs because they are now able to claim a portion of a consumer's cellular budget they have never had access to.

Voice-over-cable modem is just the first in a list of new services next-generation EMTAs will enable. Carriers have begun to plan exciting service deployments that will launch as yet unrealized revenue streams. The primary challenge that designers face is that the requirements of the VoCM market will probably change before they can even ship their current design. Already there is a strong push to provide more than just the minimum solution.

Convergence visions aside, what matters most when you consider the viability of a product is its overall cost. A flexible architecture with the right level of integration - one that allows you to optimize performance through a product family as well as your next-generation design - is essential to enabling future expandability without cost-burdening your current design. In this way, you can successfully balance the feature requirements of today with the as yet unknown demands of tomorrow.