

# Developing Video Phones with ARM Processor-based Solutions

**Author:**

Michael Ward, Trinity Convergence

**Synopsis:**

*Applications processors have evolved to the point where it is now possible to move the voice processing from the DSP to the ARM processor-based applications processor. Using such an approach provides numerous advantages, including cost savings, increased battery life for portable devices, and reduced footprint for the device. This article discusses the use of ARM processor-based applications in a video and voice over-IP (V2IP) enabled platform and other topics.*

Video telephony has been a promise for more than 40 years. The AT&T Picturephone, introduced at the 1964 World's Fair, delivered voice and video at \$16 to \$24 per minute in 1964 dollars (\$78 - \$118 in today's inflation-adjusted dollars). At such price levels, it is not surprising that the technology was not rapidly adopted by the masses. Time and time again, the industry's attempts to re-introduce solutions for mass-market video telephony were faced with such significant quality challenges it could not generate more than novelty interest in the devices. Coupled with the high price of the equipment and associated service, video-phones even failed to be accepted by the text-book "early adopter". The result — video phones that residential consumers left sitting on the shelf.

This time around, things are different. One by one, the hurdles facing the deployment of a high-quality video telephony solution attainable by the average consumer are being solved. In large part, the technological progress has come on two different fronts: increasingly sophisticated voice-over-IP (VoIP) solutions, coupled with significant improvements in video compression technology such as H.264 and MPEG-4. In addition, the widespread adoption of Broadband Internet services, now approaching 50% penetration in much of North America and parts of Europe, has resulted in more than adequate bandwidth to the home. And, of course, Moore's Law continues to work its magic, delivering processing devices capable of very cost-efficient full-motion video.

The result is that by utilizing soft-VoIP device design techniques, in combination with hardware-accelerated video processing, phones can now be created that meet the aggressive cost-points needed for video telephony to finally reach mass-market potential.

For video phones to be adopted by the masses, the price for the video phone must reach the right level. The "magic" consumer price point of \$99 is the target for many providers to the residential video telephony market. While this is a challenging price point for the current generation of technology, many VoIP service providers are subsidizing video phones in exchange for a year or more commitment to their service. With current video phone models retailing for \$800 or more before service provider subsidies, a heavy burden is placed on the service provider, and can reduce the service provider's desire to offer such services.

The price-point of current VoIP-enabled video phone designs is driven by the high number of specialized components they require. Traditional video- and voice-over-IP (V2IP) designs often contain as many as three independent processors — a DSP for voice processing, a video encode and decode processor (either a dedicated video processor or a generalized DSP handling video encoding and decoding), and an applications processor handling VoIP call signaling, the user interface and generalized control of the device. These processors, coupled with other necessary electronic components, such as the camera, LCD, and audio codec, quickly drive the overall electronics Bill of Materials (eBOM) to the \$500 realm, before any costs for software, packaging or manufacturing are added.

Fortunately, technology is emerging that allows the eBOM for a video phone to be cut at least in half, if not more. As a provider of VoIP solutions for both infrastructure equipment and consumer devices, Trinity Convergence has taken a unique approach in how it helps OEMs implement voice/video applications — an approach that is proving to significantly lower the overall eBOM. Trinity's video



and voice over IP solution for consumer devices (VeriCall Edge™) is utilizing highly integrated applications processors and DSP-free techniques for VoIP processing. The resulting benefit is that the multiple processors previously required can be combined into a single device, yielding significant savings in costs, power and device size.

The remainder of this article will take a close look at how VoIP services can be supported by general-purpose applications processors based on ARM9E and ARM11 processor architectures.

### Audio Processing

In traditional video phone devices, a DSP handles the packet voice processing (voice encode/decode, tone generation and detection, echo cancellation, noise reduction, etc.), a separate DSP or dedicated co-processor handles the video encode & decode, and an applications processor manages the VoIP call control protocol and user interface (Figure 1).

In this multi-device architecture, tasks are relegated to different components of the system for processing. As such, the task of coordinating and managing the overall system is increased. The increased compo-

nent count increases the size of the design and the overall complexity of the circuit board. Given the multiple devices, each with varying power requirements, the power supply design within the system may become very complex, requiring the use of multiple voltage converters to generate the various power rails needed.

The traditional architecture utilizing separate processors for the voice, video and system control requires multiple programming models and development tool chains. This approach results in the need for larger development teams, increased training requirements and additional costs for the development tools.

and increased battery life, or additional processing power for the remainder of the application. To put this in context, a typical G.729AB codec optimized with the DSP extension instructions can save

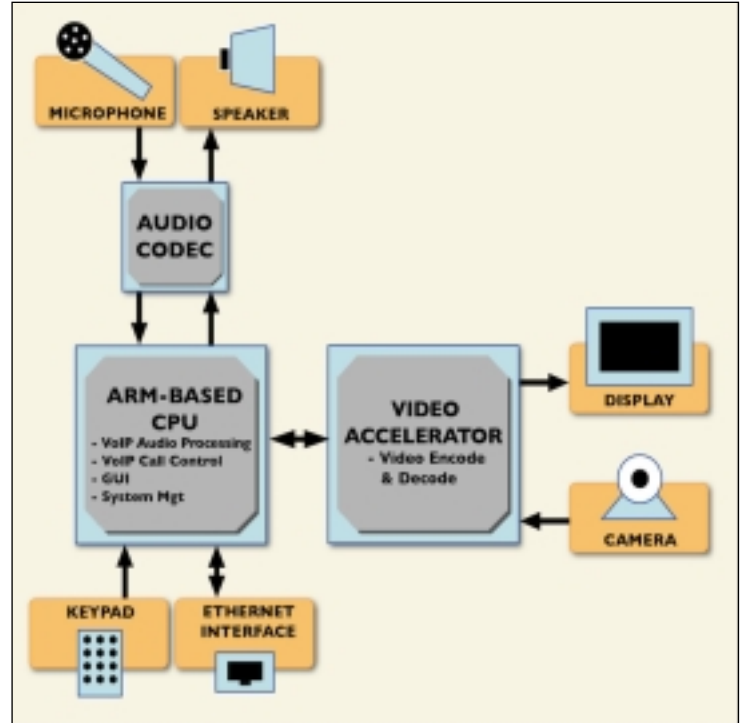


Figure 2: Architecture of a Soft-VoIP based Video Phone

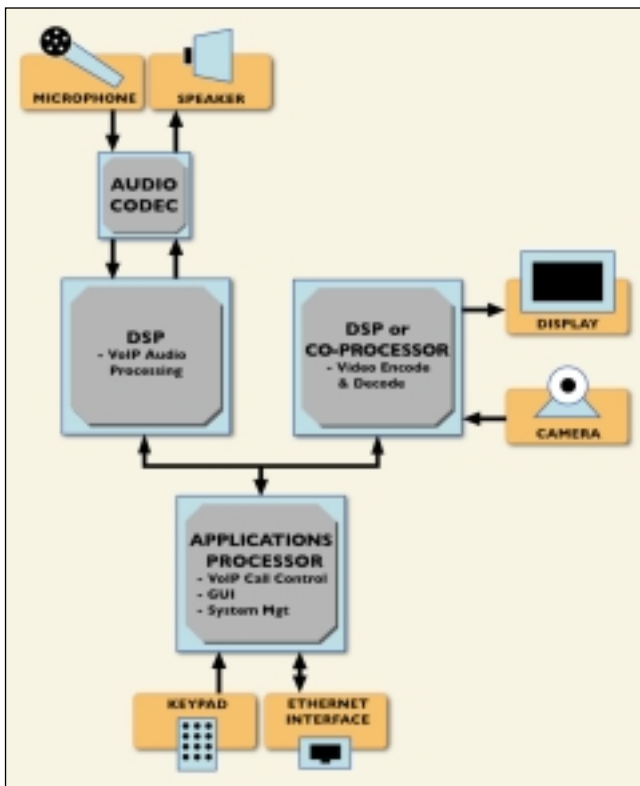


Figure 1: Architecture of a traditional Video Phone

General-purpose applications processors based on ARM9E and ARM11 processor architectures have increased in processing power, to the point where it is possible to move the audio processing tasks usually performed by a DSP to the applications processor.

The ARM9E and ARM11 instruction set extensions are highly flexible and can be used to optimize virtually any type of media processing. Using the 'E' extensions of the ARM926EJ-S reduces the Mhz required to execute a typical voice encoder by as much as 20%, when compared to its execution on an ARM9 family processor without the 'E' extensions. This saving in Mhz translates to either a lower clock speed

approximately 5 Mhz over a well optimized implementation for an architecture without these instructions. However, this gain is magnified when implementing a more complex, wide-band codec such as G.722.2, where approximately 20 Mhz can be saved through a DSP extension based implementation.

VoIP codecs (G.711, G.729AB, G.723.1, iLBC), audio processing (DTMF and call progress tone detection / generation), voice quality enhancement (line and acoustic echo cancellation, jitter buffers, etc.) and other similar functions can now all be effectively executed on the applications processor if carefully implemented with assembly coded and hand-optimized software, while utilizing some form of hardware acceleration for the video encode and decode (Figure 2).

Detailed knowledge of both the processor architecture and VoIP audio processing algorithms is needed to effectively implement this processor intensive functionality. While a significant integration effort is



required to port these voice processing modules to the appropriate applications processor architecture, the resulting benefit is well worth the investment.

Due to the real-time nature of VoIP traffic, and the need to provide a wide-range of VoIP codec support to ensure interoperability across devices, OEMs should take care to ensure that they are implementing a flexible VoIP framework. Such a framework needs to provide run-time selection and configuration of the appropriate VoIP codec, as well as dynamic configuration of the media processing elements, within a given media channel. The framework, and its associated scheduler component, must ensure that all algorithms required for a given channel definition are executed in the time period allowed. While in a single-channel system, the task of scheduling these algorithms is little more than a series of consecutive calls to the appropriate algorithms in order, multi-channel systems offer a more complex scenario in which different VoIP codecs may be required for each channel, as well as certain channels requiring echo cancellation while others do not. Videophones typically are "single-channel" systems, although capable of 3-way audio calling with the audio mixed locally on the videophone, so the need for multi-channel support arises.

Simplifying life for device designers, implementations of VoIP frameworks such as

VeriCall Edge provide a highly optimized solution in a form that can be quickly integrated into the end product design. By providing all the requisite media processing algorithms as well as SIP or H.323-based VoIP call control combined in a flexible framework integrated for ARM9, ARM9E and ARM11 based devices, the videophone developer can focus on support for value-added services on the device.

By merging the packet voice processing onto a general-purpose processor, only one device is required to handle the VoIP-related tasks of the video phone. The VoIP call control, user interface and general telephony control application all execute on a general-purpose applications processor. As these other tasks do not exhibit the same time-critical attributes, they are mainly implemented in 'C', without the need for optimized assembly code. Merging the media processing and call control / system management onto a single applications processor enables a simplified design which reduces the device count, cost, size. In addition, it also removes the need for separate development efforts and separate development tool chains for the VoIP media processing and system control.

To further increase the level of system integration, and gain additional benefits of lower-power, smaller foot print and overall

lower system costs, the use of a System-on-Chip (SoC) that contains the general-purpose processor along with a dedicated video co-processor or DSP can be used (Figure 3). In a typical implementation with VeriCall Edge, the device can take advantage of such an architecture by running all the VoIP media processing on the ARM core while the software framework configures and manages the hardware based video-acceleration block to control the video stream.

### Video Processing

The most processing-intensive task within a video phone is the processing of the video itself. The specific processing required may vary considerably depending on the size of the transmitted and received image, as well as on the particular encoding scheme used. H.263, H.264 and MPEG-4 are among the most commonly used codecs for video telephony. H.264 has the benefit of requiring less bandwidth to transmit a video image of quality comparable to that of H.263, but requires significantly more processing power (typically requiring significantly more costly processors) to achieve this higher level of compression. The specific video compression scheme, desired maximum frame rate and range of supported resolutions will drive the "right" solution for video processing within a device.

While it is possible to provide a certain level of lower frame rate (10 - 15 fps) video at lower resolutions (SQCIF to QCIF) using software codecs running on the ARM core, the processing required for full-motion video (25 - 30 fps), using complex encoding techniques such as H.264 requires a dedicated video encode / decode accelerator. Video accelerators are discrete processing blocks included in as a part of the overall SoC design specifically designed to execute media processing. Video data is transferred from the ARM core to the video accelerator block where the video encode or decode task is carried out. By off-loading this intensive processing to a dedicated hardware component, the ARM core is able to efficiently process the VoIP audio, while controlling the video accelerator to provide synchronized audio and video for a high-quality experience.

### See the Future...

The enabling technology is now available to deliver video telephony to the residential market. Success, however, will require equipment manufacturers of both infrastructure equipment and consumer devices to innovate on their VoIP implementations. By coupling ARM-based processors with soft-DSP VoIP techniques and hardware-accelerated video processing, video phones can be developed at a savings of as much as 50 percent in the eBOM of the device. This dramatic savings in the product cost will allow video phones to be marketed at price-points that finally enable mass-market adoption.

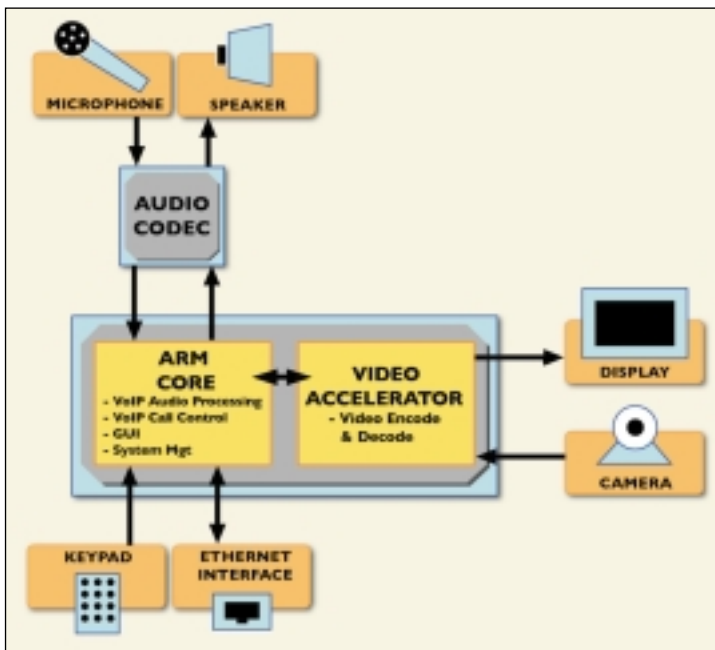


Figure 3: System on Chip Architecture with Soft-VoIP Approach for Video & Voice-over-IP