White Paper

Designing High-Fidelity Videophones on the Freescale i.MX27 Multimedia Applications Processor

Andy Lilly and David Brown
Trinity Convergence

Document Number: IMX27MMAPPRCSRWP
Rev 0
1/2008

Overview

All of the fundamentals are now in place to make videophones—wired and Wi-Fi—available to the consumer market.

Thanks to competition amongst cable and telephone service providers—and a handful of WiMax upstarts—the aggressive rollout of broadband networks, coupled with highly-capable video/voice processing technology companies such as Freescale Semiconductor are overcoming the technical challenges of bringing video communication into the home and office.

Price and performance—both audio and video—are usually perceived to be the key hurdles in delivering a mass market high-fidelity solution. This paper addresses those concerns in describing how the Freescale i.MX27 multimedia applications processor, running VeriCall® video and voice over IP (V2IP) software from Trinity Convergence, can be used to rapidly and cost-effectively design high-fidelity videophones.

Contents

1 Design Considerations

Designing, implementing and manufacturing a video and voice over IP (V2IP) phone requires a very capable processor and significant resources for software development, integration and validation. To build a compelling solution the following four key elements need to be brought together:

- Silicon platform
- Embedded video/voice over IP framework
- Application services layer
- Graphical user interface (GUI)

1.1 Silicon Platform

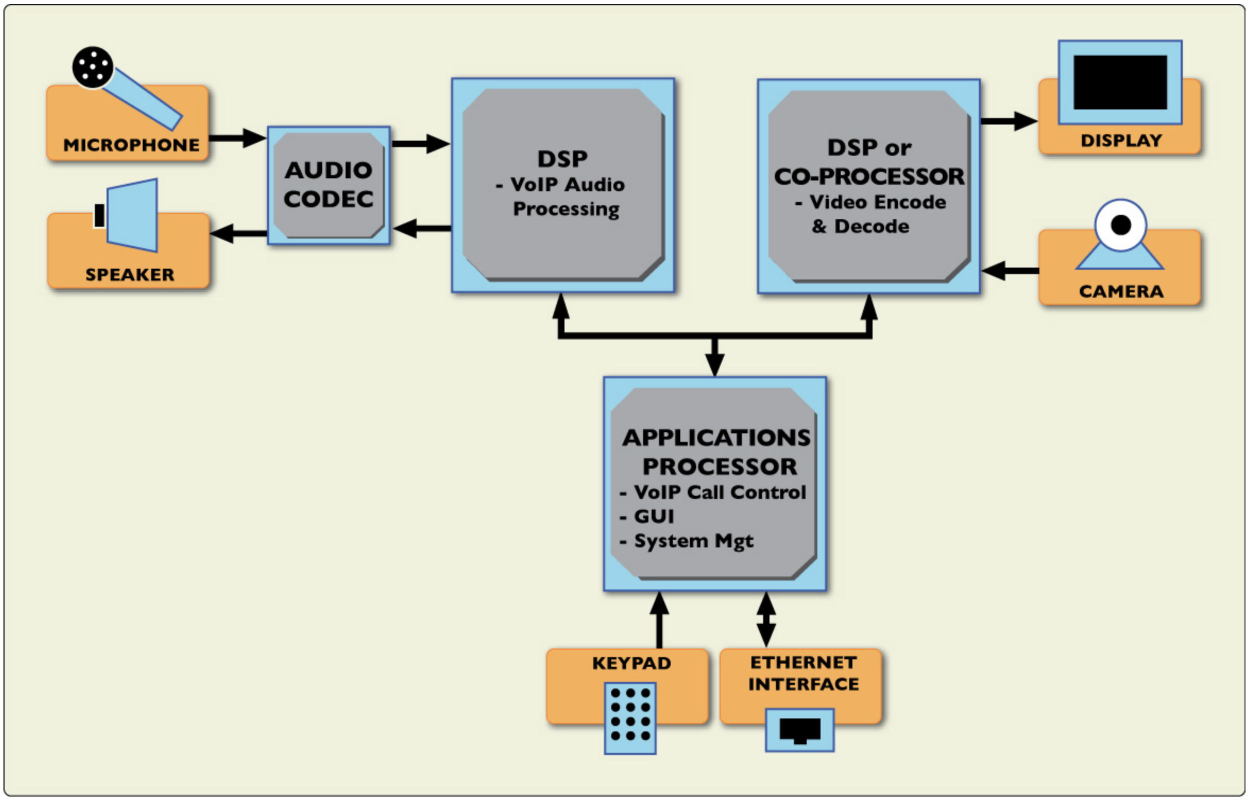
The i.MX27 multimedia applications processor from Freescale has helped to make consumer video telephony a much more viable consumer product offering. As the first i.MX family processor capable of delivering two-way H.264 video in VGA resolution, Freescale is delivering a very compelling solution—one capable of helping ODMs and OEMs reach mass market price points for V2IP-enabled consumer electronics. Complete with a video accelerator hardware block, the i.MX27 processor's ability to output high-quality, low bit-rate H.264 video meets a key requirement for enabling delivery to consumer homes. And, in common with other Freescale i.MX multimedia processors, the i.MX27 processor also offers a variety of power saving techniques to give extended life when used in battery-powered portable equipment. We'll explore this in more detail.

The architectures used by first generation videophones tended to use separate processors for the voice, video and system control functions. Due to the processing requirements these products typically used processors optimized for intensive media processing operations (i.e., digital signal processors, or DSPs). Figure 1 shows an example of such a phone, comprising:

- One DSP to handle voice processing functions, including voice encode/decode; tone generation and detection; echo cancellation and noise reduction
- One DSP or dedicated co-processor to handle the video encode/decode
- One applications processor managing the VoIP call control protocol and user interface

This approach dramatically increases the complexity, cost and power requirements of the hardware. It also requires multiple programming models and development tool chains which, in turn, demands larger development teams, a broader skill base and longer integration period. The cumulative result of these hardware and software factors is an expensive bill of materials and a very high development cost.

Figure 1: First generation videophones requiring three separate processors



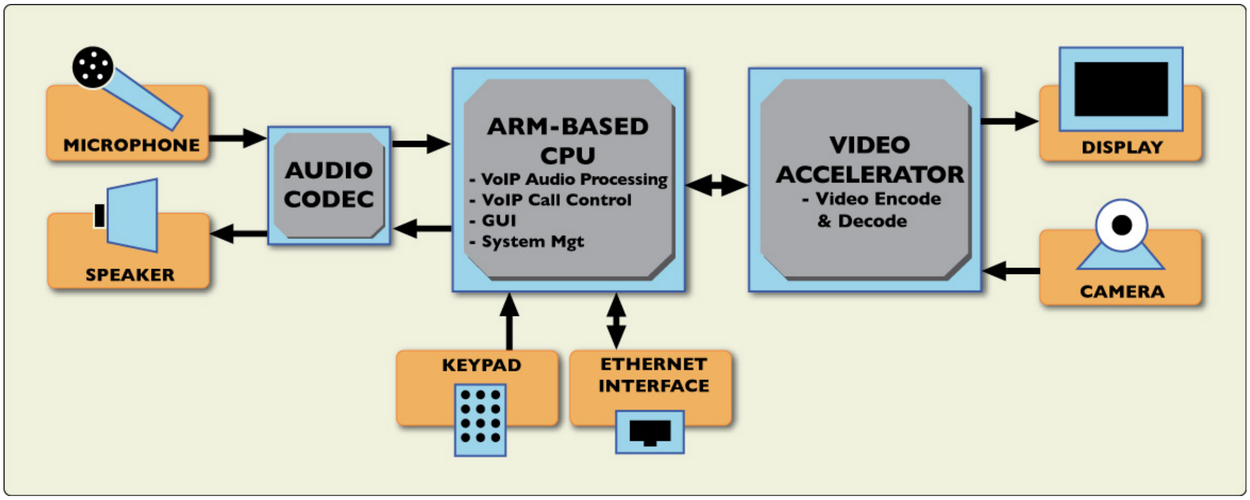
Since the first generation of IP videophones were introduced, general-purpose applications processors have increased in processing power to the point where it is possible to move all of the audio processing tasks usually performed on the audio DSP to the applications processor.

This isa first step in simplifying the hardware design and reducing its cost and power consumption. This is of particular benefit to the Wi-Fi videophone market because of the resulting extended battery life for hand-held devices.

VoIP codecs (such as G.711, G.729AB, G.723.1, Speex, iLBC, G.722.2), 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 appropriate use of hand-optimized assembly code for the most processing-intensive functions.

The second step forward is including dedicated hardware acceleration for the video encode/decode process—now found on an increasing number of applications processor families. For example, the i.MX21 multimedia applications processor provided commercially viable H.263 or MPEG4 video up to CIF resolution—useful for screens of 3-4” across. Its “big brother,” the i.MX27 processor, provides high quality full-duplex error-tolerant H.264 video at up to 30 frames per second (FPS), providing excellent video communications on much larger LCD screens, up to 8-10” across, or on a television. Such processors lead to the ideal videophone architecture shown in Figure 2.

Figure 2: A new paradigm for designing videophones



It must be remembered that video functions need to encompass more than just encode/decode to be of use in a product—typically other video operations are required (such as scaling, resizing, mirroring, rotation and color space conversion), as well as support for multiple graphics layers (enabling operations such as picture-in-picture or menu overlays on the video image). Efficient (direct) interfaces to the camera and screen give optimal performance and standard programming interfaces (e.g. Video for Linux 2 or "V4L2") are highly desirable to simplify integration. The i.MX27 processor includes the enhanced Multimedia Accelerator (eMMA) which delivers these necessary capabilities through its V4L2 drivers.

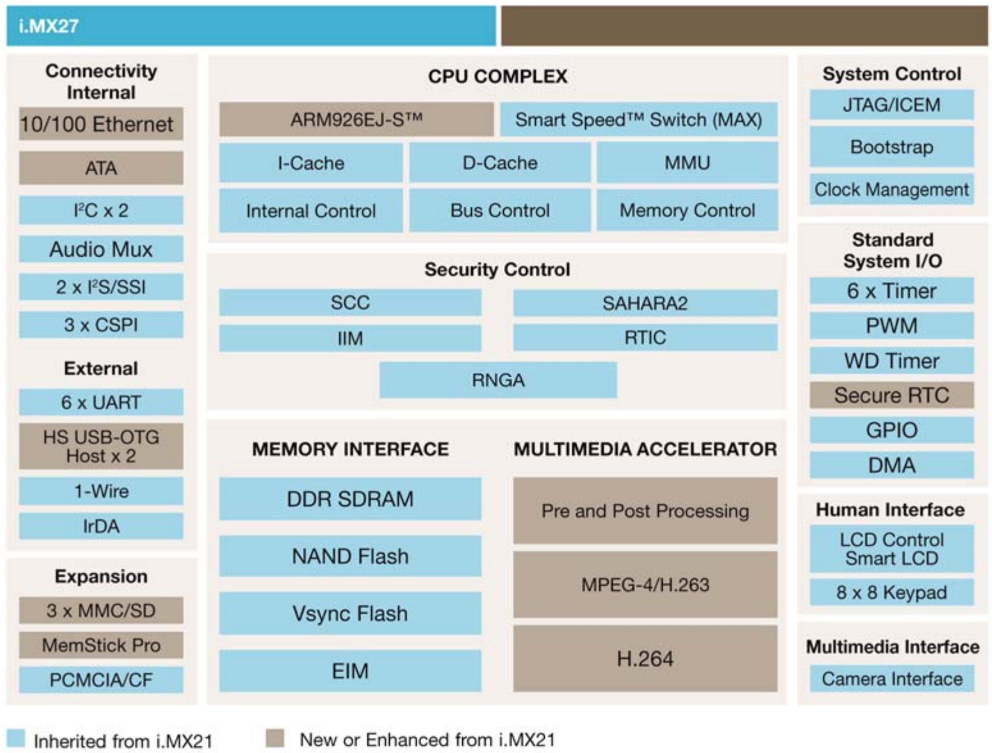
An additional benefit of the increased processing capabilities on today's applications processors is that we can use advanced operating environments such as embedded Linux® or Windows® CE operating systems to effectively partition the control and media processing required in V2IP systems. In turn, this leads to simplified software development using a single processor and tool chain, reduced integration and validation effort and—of benefit to everyone from design to the end customer—reduced time to market and lower cost at every stage of development.

Video Processing

Videophones use one or more of the following video compression algorithms: H.263, H.264 or MPEG-4. Of these, H.264 (also known as MPEG-4 AVC – advanced video coding) is the most sophisticated in its ability to deliver low bit-rate, high-quality video in real-time. Compared to H.263 however, this more advanced codec requires significantly more processing power. Only the most recent processors, such as Freescale's i.MX27 processor, have the capability to make H.264 both technically and commercially viable for mass market videophones. The i.MX27 video codecs even handle two video streams, enabling multi-way video calling.

The full range of capabilities required of such an applications processor is readily available in the i.MX27 processor, as shown in Figure 3. In this modern world where information security is increasingly important to the consumer, the inclusion of hardware acceleration features, such as the encryption algorithms provided in the SAHARA2 block, offer a differentiated set of services relevant across a variety of communication devices.

Figure 3: Freescale i.MX27 applications processor architecture for videophone products



2 Embedded Video/Voice over IP Framework

At the heart of a V2IP design is the embedded voice and video processing, and the software elements that control and manage the data flow through the system (the framework). OEMs and original design manufacturers (ODMs) have three options for developing the V2IP framework:

1. Build a complete V2IP software framework from the ground up.
2. License components and software stacks and provide the integration, validation, silicon porting and interoperability testing.
3. License a pre-integrated and proven framework from a third party.

The expert resources and timescales required to undertake Option 1 are prohibitive for all but the largest development organizations. Option 2 reduces both these factors to some degree but still places a heavy integration burden on the developer, in addition to increased risk where the bought-in components come from different vendors.

Unless IP and networking software development are core strengths of your organization, one of the fastest, lowest risk and most cost-effective option is to license a third-party framework already optimized for execution on your target processor. This allows you to focus on the areas such as the user interface and branding that will bring most value and differentiation to your new product.

A highly optimized solution can be quickly integrated into the end product design. Such a solution must provide all of the media processing algorithms and V2IP call control, combined in a flexible framework allowing you, the end product developer, to focus on designing a capable, value-added device.

Given the real-time nature of IP traffic, a tightly integrated V2IP framework is critical for ensuring reliable and stable voice and video communications. From a comprehensive media processing library to a range of quality of service (QoS) and networking clients, a V2IP software framework ultimately determines the quality and performance of the video/voice communications.

It is important that the V2IP framework is truly flexible. It should provide run-time selection and configuration of the appropriate voice and video codecs, as well as dynamic configuration of all aspects of the media processing elements, within a given media channel. It must also be able to manage the runtime bandwidth of each channel, balancing audio and video quality against the constraints of the user's network capacity. The framework, and its associated scheduler component, must ensure that all algorithms required by a given channel are executed within the time period allowed by the real-time system. 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 voice and video configurations (from codecs to echo cancellation mechanisms) may be required for each channel. Early videophones are typically single-channel systems, but future models will generally be capable of three-way audio/video calling.

Designing a VoIP or videophone today, let alone a Wi-Fi videophone, requires product differentiation and support for next generation services and functions. Legacy VoIP phones provide "toll quality" voice codecs (such as G.711) and current widely available videophones use basic video compression functionality (e.g. H.263). These baseline capabilities can support a personal video conference and have been used successfully for a number of years.

However, in today's high-fidelity and high-definition world, next generation videophones must support wideband audio and advanced video compression technologies if they are to be embraced by the masses. Technologies such as AMR-WB (G.722.2) audio and H.264 video compression significantly enhance the communications experience, providing more pristine video images at lower bit rates and much sharper and more natural-sounding audio between the two parties. The requirement for hands-free audio—which is the preferred means of using most hand-held or desk videophones—introduces the additional processing load of an acoustic echo canceller (AEC). The combination of these advanced communications technologies calls for a processor that, in addition to hardware acceleration of the video, has the power to fulfill the user's expectations. Where a standard quality phone might have required a processor running at 200 MHz, a high fidelity videophone needs closer to 400 MHz—a level of processing power which can be obtained by using Freescale's i.MX27 multimedia applications processor.

In addition to wideband audio and higher definition video, there are a variety of technologies that benefit the end user by enhancing the reliability, performance and quality of V2IP communication. Going forward, we believe that the features outlined in Table 1 will be required for competitive VoIP and V2IP solutions.

Table 1: Features and Benefits of Competitive VoIP and V2IP Solutions

Audio Capability	**Customer Value**
Wide variety of audio codecs (e.g. G.711, G.723.1, G.726, G.729AB, iLBC)	Maximize interoperability with both legacy and future phones
Wideband audio codecs (e.g. G.722.2, Speex wideband, G.722)	Enhanced high fidelity user audio experience
Audio playback and record	Record and play back conversations
3-way calling (local audio mixing)	Supports multiple participants in a call

Audio Capability	Customer Value
Music playback (MP3, SP-MIDI)	Play music, audio clips, ring tones, etc.
Streaming audio/video (RTSP)	Stream music, videos, radio shows, etc. without the need to first download to the device
Acoustic echo cancellation (full duplex, wideband)	Natural hands-free conversation
Line echo cancellation (G.168)	Removes audio echo when used with legacy telephony equipment
Country-specific call progress tone generation / detection	Built in support for widely recognized call tones wherever your customer is found
Universal tone generator	Ability to customize call tones, ring tones, phone alerts, etc.
Gain control – automatic and manual modes	Copes with variability in the audio environment
DTMF detection / generation / relay	Tone dialing and interactive tone controls
Up/down sampling for 8, 16, and 44.1 kHz	Support for low and high fidelity audio channels using standard audio interface
Video Capability	**Customer Value**
Standard video codecs (e.g. H.263, MPEG-4)	Interoperability with existing videophones
Enhanced video codecs (H.264 / MPEG-4 AVC)	High quality, high resolution, lower bit rate video communication
Video playback and record	Record and playback video conversations and video clips
Networking Capability	**Customer Value**
Network Address Translation (NAT) traversal (STUN, TURN)	Necessary for communications across all types of networks and gateways

3 Seamless Integration of Applications and Graphical User Interface

Once the system designer has selected a capable framework to handle voice processing, video processing, call setup and NAT traversal, the focus moves to differentiating the product from other V2IP devices on the market by designing and building the user experience.

Users now expect a far superior experience than they did just a few years ago, combining many factors ranging from the quality of the key components used in the videophone (e.g. speaker, microphone, camera and display) to the look and feel of the user interface (a more subjective measure, and therefore less easy to analyze and perfect).

With better display technology on virtually all the devices being used for real-time, personal communications, the graphical user interface (GUI) becomes an increasingly important part of the user experience. Even the most basic Wi-Fi VoIP phones today offer full color display GUIs with features such as animated menus, photo caller display and instant messaging.

The integration of a GUI with an embedded V2IP framework is non-trivial. The largest hurdle facing most developers is that the types of processing in a GUI and a V2IP framework are inherently different:

- **V2IP framework**: highly responsive, media-oriented, real-time processing
- **GUI**: reactive, user-oriented, event-driven processing

A well designed V2IP framework will offer a capable application programming interface (API) that requires minimum interaction with the GUI. Specifically, the API should only require invocation in response to events generated by the user or the network. This split avoids the uncomfortable union of event-driven and real-time-media processing elements, enabling a simple integration that allows the developer to focus on an intuitive, value added GUI.

Trinity Convergence has recognized the dichotomy here and delivers dual software solutions: VeriCall Edge provides real-time management of the embedded V2IP framework layer, while VeriCall Phone Editions™ builds upon this foundation by adding a complete, customizable graphical user interface.

Understanding your own organization's strengths and weaknesses is the key consideration in effectively managing this part of the development process and determining the most productive methodology for taking a product to market. In recognition of the fact that implementing a V2IP framework is an immense undertaking, VeriCall Edge software is specifically designed to deliver rapid return-on-investment by handling all the complexities of the real-time V2IP functionality, from SIP messaging to multi-user video handling. At the same time, VeriCall Edge retains an extensive set of APIs that still allow your user application to access every aspect of the built-in feature set. Should you have the appropriate resources in-house to build a user interface from scratch for your product, then VeriCall Edge on the i.MX27 multimedia applications processor is your ideal solution, delivering a pre-integrated, fully validated V2IP solution.

Operating Systems

VeriCall® software and the i.MX27 multimedia applications processor offer OEMs and ODMs the ability to develop video and voice over IP devices on multiple real-time operating systems, including Microsoft Windows CE and embedded Linux.

While many of today's current VoIP devices are running on embedded Linux, the market is showing increased interest in leveraging the now more mature Windows® CE platform. This is especially true when it comes to multimedia devices such as PDAs and PMPs where OEMs are looking at adding IP communications functionality into existing products.

Whether you decide on embedded Linux or Windows CE, VeriCall software and Freescale i.MX processors offer a fast time-to-market solution that is based on industry standards such as SIP and H.264 video. In addition, the third party ecosystem for both operating systems is extensive and offers solutions for board support packages (BSPs), add-on software modules and hardware reference designs.

For more information on solution partners working with Freescale Semiconductor and Trinity Convergence, visit www.freescale.com/fwdn and www.trinityconvergence.com.

However, even a phone application and GUI may be time-consuming to produce. The basic phone features must be created before any beneficial product value or brand differentiation can be added. In order to hit this value-add phase as quickly as possible, VeriCall Phone Editions on the i.MX27 processor delivers a fully featured phone application and GUI, leaving your key developers to add the finishing touches (e.g. changing the appearance or "skin" of the user interface) for a ready-made product. In addition, there is a clear technical and financial benefit to using an application that has been built to seamlessly integrate with the V2IP framework, with a single source supporting all aspects of your product solution.

4 Conclusion

The bandwidth, processing power and software systems are now available to develop robust and reliable voice and video products for both fixed and Wi-Fi-enabled networks.

The success of videophone manufacturers will be driven by their ability to deliver innovative products to the marketplace that are easy to use, reliable and meet the right price points. The most profitable way to achieve these targets is to ensure the tight integration and interoperability of the critical media processing, network management and user interface/application software, i.e. basing the product on a well-validated underlying V2IP framework, running on a suitably powerful applications processor.

As demonstrated in this paper, Freescale's i.MX27 multimedia applications processor provides a solid foundation on which to build a broad range of V2IP products. Rapid delivery of a high fidelity audio, premium quality video product is enabled by the VeriCall Edge and VeriCall Phone Editions software from Trinity Convergence, allowing you to beat your competitor time-to-market and maximize the return on your development investment.

We would like to get your feedback on this paper and address any questions regarding the challenges and recommendations described above. Please email your comments and feedback with a subject line of "Videophone Article" to info@trinityconvergence.com and we'll get back to you faster than you can say "low latency."

How to Reach Us:

Home Page:
www.freescale.com

Web Support:
http://www.freescale.com/support

USA/Europe or Locations Not Listed:
Freescale Semiconductor, Inc.
Technical Information Center, EL516
2100 East Elliot Road
Tempe, Arizona 85284
+1-800-521-6274 or +1-480-768-2130
www.freescale.com/support

Europe, Middle East and Africa:
Freescale Halbleiter Deutschland GmbH
Technical Information Center
Schatzbogen 7
81829 Muenchen, Germany
+44 1296 380 456 (English)
+46 8 52200080 (English)
+49 89 92103 559 (German)
+33 1 69 35 48 48 (French)
www.freescale.com/support

Japan:
Freescale Semiconductor Japan Ltd.
Headquarters
ARCO Tower 15F
1-8-1, Shimo-Meguro, Meguro-ku,
Tokyo 153-0064, Japan
0120 191014 or +81 3 5437 9125
support.japan@freescale.com

Asia/Pacific:
Freescale Semiconductor Hong Kong Ltd
Technical Information Center
2 Dai King Street
Tai Po Industrial Estate
Tai Po, N.T., Hong Kong
+800 2666 8080
support.asia@freescale.com

For Literature Requests Only:
Freescale Semiconductor
Literature Distribution Center
P.O. Box 5405
Denver, Colorado 80217
1-800-441-2447 or 303-675-2140
Fax: 303-675-2150
LDCForFreescaleSemiconductor@hibbertgroup.com

Document Number: IMX27MMAPPRCSRWP
REV 0