



WP328 (v1.0.1) November 21, 2007

# *FPGAs Driving Voice-Data Convergence*

*By: Amit Dhir*

---

Over the last few years, data exchange between people using the Internet has gained popularity, with over 6.9 trillion emails being exchanged last year. There is a big push to use the same backbone for voice traffic. This article gives an overview of voice data convergence technologies, the benefits to the users and some of the significant challenges facing the designers of these systems.

---

© 2001-2007 Xilinx, Inc. All rights reserved. All Xilinx trademarks, registered trademarks, patents, and further disclaimers are as listed at <http://www.xilinx.com/legal.htm>. All other trademarks and registered trademarks are the property of their respective owners. All specifications are subject to change without notice.

NOTICE OF DISCLAIMER: Xilinx is providing this design, code, or information "as is." By providing the design, code, or information as one possible implementation of this feature, application, or standard, Xilinx makes no representation that this implementation is free from any claims of infringement. You are responsible for obtaining any rights you may require for your implementation. Xilinx expressly disclaims any warranty whatsoever with respect to the adequacy of the implementation, including but not limited to any warranties or representations that this implementation is free from claims of infringement and any implied warranties of merchantability or fitness for a particular purpose.

## Voice-Data Convergence (Voice over Internet Protocol)

Convergence in networking refers to the ability to transfer data and voice (and/or video) traffic on a single network. Voice over IP, also known as IP telephony, and packet-voice, is the transmission of voice traffic in packets using the Internet Protocol (Internet backbone).

Current voice telephony is based on a circuit-switched infrastructure and uses the PSTN (public switched telephone network). When a call is placed, the PSTN reserves 64Kbps, end-to-end bandwidth for the duration of the call, on a fixed channel. A voice call generally does not utilize the full channel bandwidth. While, PSTN supports full duplex transfer, phone calls involve one person talking and the other listening and vice versa. There are many periods of silence where the network transmits no information, and hence wastes network bandwidth.

In VoIP networks, the packetization of voice happens in real-time. VoIP also decreases the bandwidth utilized significantly, since multiple packets can be transmitted simultaneously. The SS7 and TCP/IP networks are used together to setup and tear down the calls, along with Address Resolution Protocol (ARP).

Process of creating IP packets:

**Step 1:** An analog voice signal is converted to a linear pulse code modulation (PCM) digital stream (16 bits every 125msec).

**Step 2:** The line echo is removed from the PCM stream and is further analyzed for silence suppression and tone detection.

**Step 3:** The resulting PCM samples are converted to voice frames and a vocoder compresses the frames. G.729a creates a 10ms long frame with 10Bytes of speech. It compresses the 128kbps linear PCM stream to 8kbps.

**Step 4:** The voice frames are integrated into voice packets. First, a RTP packet with a 12-byte header is created. Then an 8-byte UDP packet with the source and destination address is added. Finally, a 20-byte IP header containing source and destination IP addresses is added.

**Step 5:** The packet is sent through the Internet where routers and switches examine the destination address, and route and deliver the packet appropriately to the destination. IP routing may require jumping between networks and pass through several nodes.

**Step 6:** When the destination receives the packet, the packet goes through the reverse process for playback.

The IP packets are numbered as they are created and sent to the destination address. The receiving end must reassemble the packets in their correct order (when they arrive out of order) to create voice. The IP addresses and telephone numbers must be mapped properly.

Description	PSTN	Internet
Designed for	Voice only	Packetized data, voice & video
Bandwidth Assignment	64Kbps (dedicated)	Full-line bandwidth over a period of time
Delivery	Guaranteed	Not guaranteed
Delay	5-40ms (distance-dependent)	Not predictable (usually more than PSTN)
Cost for the Service	Per-minute charges: long distance Monthly flat rate: local access	Monthly flat rate for access
Voice Quality	Toil quality	Depends on customer equipment
Connection Type	Telephone, PBX, switches with frame relay and ATM backbone	Modem, T1/E1, Gateway, Switches, ISDN, bridges, Routers, Backbone
Quality of Service	Real-time delivery	Not real-time delivery
Network Management	Homogeneous and interoperable at network and user level	Various styles with interoperability established at network layer only
Network Characteristics (Hardware)	Switching systems for assigned bandwidth	Routers & bridges for layer 3 and 2 switching
Network Characteristics (Software)	Homogeneous	Various interoperable software systems
Access Points	Telephones, PBX, PABX, ISDN, switches, high-speed trunks	Modem, ISDN, T1/E1 Gateway, high-speed DSL/cable modems

Figure 1: PSTN (circuit-switched) vs. IP (packet-switched) networks

## Motivation and Market

The integration of voice, video and data allows the use of a unified packet network, and thus reduces bandwidth consumption by 8:1 in favor of packet-based networks. By eliminating the voice infrastructure, the costs of maintaining both networks are eliminated. Web users are demanding free voice and video communications, as voice is the logical step from ubiquitous Internet mail and instant messengers. VoIP also provides enhanced features like flexible call routing and networked multimedia applications. The ability to use voice and video as part of the Web experience helps sell and support consumers, and improves site stickiness in portals, communities, directories and audio ads. Corporations can reduce costs using VoIP services for distance learning, customer support and remote sales presentations. Growing digital convergence and networking consumer devices in today's homes, requires a low-cost, integrated voice-data-video access to the Internet.

Cahners In-Stat Group projects that sales of VoIP equipment reached \$61 million in 1998 and will exceed \$3.8 billion in 2003. The VoIP market is expected to grow from 7.7 billion minutes in 2000, to 500 billion minutes by 2005, according to Probe Research. They also forecast the market for VoIP gateway equipment will increase from \$1.2 billion in 2000 to \$10 billion by 2005.

## VoIP Variations

Fax over IP (FoIP) uses the same technology as VoIP, although fax transmissions can be distributed in a non real-time fashion through the Internet. FoIP service vendors provide servers for buffering FAX traffic, multi-way distribution service, and FoIP gateways located strategically around the country.

Voice over DSL (VoDSL) transports VoIP traffic over DSL network. The VoIP gateway (maintained by a CLEC) interconnects VoIP traffic to a class 5, voice switch and the PSTN. Subscribers use IP phones to access telephone service.

Voice over Cable (VoCable) delivers Internet and voice services over the cable infrastructure. Cable companies are struggling to upgrade buried cables from half-duplex (unidirectional) to full-duplex (bi-directional) to provide premium TV and Internet services.

## Challenges in Designing IP-Based Voice Systems

Several hurdles need to be overcome before convergence becomes a reality. The challenge involves creating a single network infrastructure that can efficiently handle the requirements of two classes of traffic that have fundamentally different characteristics. Voice and video (multimedia) streams require a constant amount of bandwidth and are sensitive to delay variations in the network. The data traffic is bursty in nature and relatively insensitive to network delay. With the connection-less nature of data networks traffic competes for bandwidth on a real-time basis.

While corporate telephony (PBX) is based on proprietary designs, IP-telephony products are all based on Internet Protocol - an open standards-based evolving technology. Designers have to adhere to standards, placing a tougher load on product-validation and testing.

The voice quality for VoIP products must match the quality of circuit-switched systems. The factors effecting voice quality are line noise, echo, voice coder used and network delay. The additional features provided by a packet-switched network need to be similar to a circuit-switched network. Features such as call waiting, toll-free numbers, credit card billing, caller ID and three-way calling need to be supported by the IP network.

## Quality of Service (QoS)

Converged networks need to support QoS for the transfer of voice and video. QoS refers to a network's ability to deliver a guaranteed level of service to a user. The service level typically includes parameters such as minimum bandwidth, maximum delay, and jitter (delay variation).

QoS must be negotiated up front, before the data transfer begins, a process known as signaling. The purpose of negotiation is to give the network equipment an opportunity to determine if the required network resources are available and to reserve the required resources before guaranteeing QoS to the client.

Another contentious issue in the quest for converged networks is the appropriate layer of the protocol stack to merge the traffic. The old way to combine the traffic was at Layer 1 using separate TDM circuits for voice and data traffic. However, it is cumbersome to configure and makes inefficient use of bandwidth since there is no statistical multiplexing between separate circuits. Up until recently, the vision for voice data convergence was the use of ATM (at layer two), because of its built-in QoS features. However, ATM has a fixed cell length, which leads to added overhead. In addition, one must manage ATM and IP networks.

The most recent trend is to merge voice and data traffic at Layer 3 over IP networks. This approach takes advantage of new IP QoS features such as the RSVP and DiffServ. These technologies also take advantage of layer two QoS features.

The Internet Engineering Task Force (IETF) has developed several technologies to add QoS features to IP networks:

Resource reSerVation Protocol (RSVP) as defined in RFC 2205 is used by a host to request specific qualities of service from the network for application data streams. Routers use RSVP to communicate QoS requests to all nodes along the path of the flow, and to establish and maintain state. RSVP requests usually result in resources being reserved in each node along the data path.

Resource Allocation Protocol (RAP) is a protocol defined within the IETF, for use by RSVP capable routers to communicate with policy servers within the network. Policy servers are used to determine who will be granted network resources and which

---

requests will have priority in cases where there are insufficient network resources to satisfy all requests.

Common Open Policy Service (COPS) is defined in RFC 2748 as the base protocol for communicating policy information within the RAP framework.

Differentiated Services (DiffServ) as defined in RFCs 2474, 2475, 2597, and 2598 uses the Type of Service field within the IP header to prioritize traffic. DiffServ defines a common understanding about the use and interpretation of this field.

Real Time Protocol (RTP) is used for the transport of real-time data, including audio and video. Using the User Datagram Protocol (UDP) for transport, it is used in both media-on-demand and Internet telephony applications. RTP consists of a data and control part; the latter is called Real-time Transport Control Protocol (RTCP). The data part of RTP is a thin protocol providing timing reconstruction, loss detection, security and content identification.

Real Time Streaming Protocol (RTSP), defined in RFC 2326, is a control extension to RTP. It adds VCR functions such as rewind, fast forward and pause streaming media.

## FPGAs Driving Voice-Data Convergence - Part 2

### VoIP Products

- Enterprise and Service Provider VoIP gateways are devices deployed between a PBX and WAN access device (router) to provide call set-up, call routing and to convert voice into IP packets. They aggregate incoming VoIP traffic and route the traffic accordingly; much like a traditional Class 5 CO switch.
- VoIP routers are standard routers with voice cards that perform packetization and compression, and the router then directs the packets to their ultimate destination.
- VoIP end stations (IP Phones) include telephone handsets, a VoIP gateway and a LAN interface.

### VoIP Gateway Technologies

The VoIP gateway includes components such as:

- Hardware Components: Digital signal processor, controller, codec, analog front end, WAN interface
- Software Components:
  - Voice processing elements - Speech coders, echo cancellers, voice activity detection (VAD), comfort noise generator (CNG), telephony
  - Protocol stack related elements - H.323 and TCP/IP protocol stack
  - Voice and LAN/WAN interface management
  - RTOS

Figure 2 illustrates the functional architecture of a VoIP gateway and its three major functional blocks:

- Voice Processing includes all functions required to encode voice data samples and packetize them for transmission.
- Telephony Signaling Gateway (TSG) subsystem performs the functions for establishing, maintaining and terminating a call.
- Network interface protocols include the TCP/IP protocol suite and the OSI Layer 2 protocols including ATM, Frame Relay, or Ethernet.
- Voice Play-out buffers the packets that are received and forwards them to the voice codec for play-out. This module provides an adaptive jitter buffer and a measurement mechanism that allows buffer sizes to be adapted to the performance of the network.
- Packet Voice Protocol encapsulates the compressed voice data for transmission over the data network. Each packet includes a sequence number that allows the received packets to be delivered in the correct order. This also allows silence intervals to be reproduced properly and detection of lost packets.

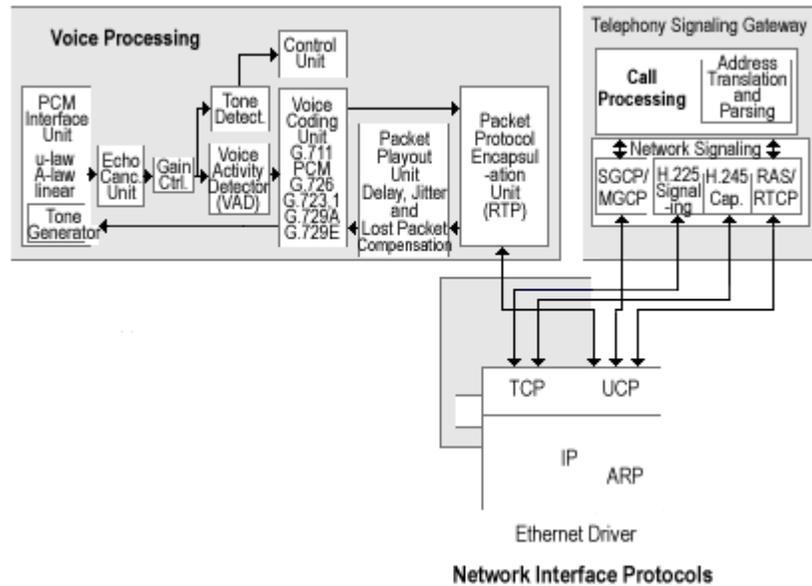


Figure 2: VoIP Gateway Architecture

## VoIP Voice Processing

Voice processing functions include the following:

- PCM Interface conditions PCM data and includes functions such as companding and resampling. This block also includes the Tone Generator, which generates DTMF tones and call progress tones.
- Echo Cancellation Unit performs echo cancellation on sampled, full-duplex voice port signals in accordance with the ITU G.165 or G.168 standard.
- Voice Activity Detector suppresses packet transmission when voice signals are not present. If no activity is detected for a period of time, the voice encoder output will not be transported across the network. Idle noise levels are also measured and reported to the destination so that "comfort noise" can be inserted into the call.
- Tone Detector detects the reception of DTMF tones and discriminates between voice and facsimile signals.
- Voice Coding Unit compresses the voice data streams for transmission. There are several different codecs used to compress voice streams in VoIP applications. Each has been targeted at a different point in the tradeoff between data rate (compression ratio), its processing requirements, the processing delay, and audio quality. Table 1 compares the various ITU codecs with respect to these parameters. The MOS (mean opinion score) parameter is a measure of audio quality rated by a subjective testing process.

Table 1: Voice Coding Standards

Standard	Description	Data Rate (Kbps)	Delay (mS)	MOS
G.711	Pulse Code Modulation (PCM)	64	0.125	4.8
G.721, G.723, G.726	Adaptive Differential PCM (ADP CM)	16, 24, 32, & 40	0.125	4.2
G.728	Low-Delay Codebook Excited Linear Predictors (LD-CELP)	16	2.5	4.2
G.729	Conjugate Structure Algebraic CELP (CS-ACELP)	8	10	4.2
G.723.1	Another CS-ACELP codec: Multi Pulse - Maximum Likelihood Quantizer (MP-MLQ)	5.3 & 6.3	30	3.5 & 3.98

- Voice Play-out buffers the packets that are received and forwards them to the voice codec for play-out. This module provides an adaptive jitter buffer and a measurement mechanism that allows buffer sizes to be adapted to the performance of the network.
- Packet Voice Protocol encapsulates the compressed voice data for transmission over the data network. Each packet includes a sequence number that allows the received packets to be delivered in the correct order. This also allows silence intervals to be reproduced properly and detection of lost packets.

## VoIP Telephony Signaling

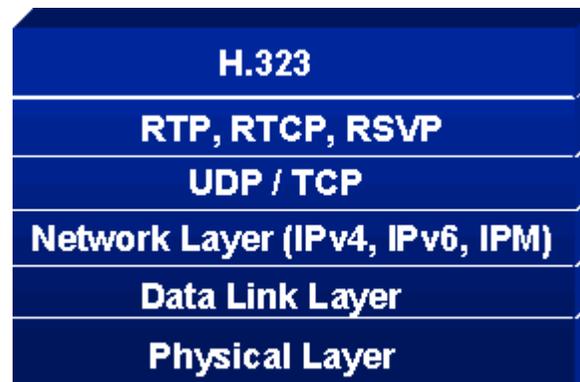


Figure 3: VoIP Protocol Structure

Telephony Signaling functions include:

**Call Processing:** Performs the state machine processing for call establishment, call maintenance and call tear down. This includes Address Translation and Parsing, which determines when a complete number has been dialed and makes this dialed number available for address translation.

**Network Signaling:** Performs signaling functions for establishment, maintenance and termination of calls over the IP network. There are two widely used standards: H.323 and SGCP/MGCP.

- H.323 Protocols H.323 is an ITU standard that describes how multimedia communications occur between user terminals, network equipment, and assorted services on local and wide area IP networks. The following H.323 standards/protocols are used in VoIP gateways:
- H.225 Call Signaling: Performs signaling for establishment and termination of call connections based on Q.931.
- H.245 Control: Provides capability negotiation between the two end-points such as voice compression algorithm to use, conferencing requests, etc.
- RAS (Registration, Admission, and Status): Used to convey the registration, admissions, bandwidth change, and status messages between IP Telephone devices and servers called Gatekeepers, to provide address translation and access control to devices.
- RTCP: Provides statistics information for monitoring the QoS of the voice call.
- SGCP/MGCP Protocols Simple Gateway Control Protocol (SGCP) is a standard that describes a master/slave protocol for establishing VoIP calls. The slave (client) side resides in the gateway (IP phone) and the master side resides in an entity referred to as a Call Agent. SGCP is adopted as part of the DOCSIS cable modem standard. SGCP is evolving to the Multimedia Gateway Control Protocol (MGCP).

Audio Applications	Video Applications	Terminal Call Manager				
G.711 G.729 G.723.1	H.261 H.263	RTCP	H.225.0 RAS	H.225.0 Call Signaling	H.245 Control Signaling	T.120 Data
RTP						
Transport Protocols & Network Interface						

## Implementing VoIP Products in FPGAs

### High Capacity VoIP Gateways

VoIP gateways support capacities of tens to hundreds of lines, but system vendors are increasing the densities to the hundreds to thousands range in anticipation of VoIP moving from the trial to adoption phase.

Creating high capacity systems is challenging due to the processing power required to handle the channels. Currently arrays of high performance DSPs are used, and the H.110 CT bus is used to transfer PCM voice streams to the line interface cards for transfer to the PSTN. Also included are DS1, DS3 or ATM port interfaces and a management processor, which typically runs the SS7 signaling software. A voice-processing card includes DSPs, memory, microprocessors (for control, signaling and data processing functions), H.110 compliant bus interface and 10/100 Ethernet interfaces. This requires a significant amount of complex glue logic including PCI bridges, memory controller and data path FIFOs. FPGAs are useful in gateway applications as:

- System-Level Glue: Implementing specialized PCI host bridges, DSP to processor interface logic, memory controllers, data path switching and FIFO functions.

- Echo Cancellation: FPGAs are more effective than DSPs at implementing functions such as high performance FIR filters and correlators.
- Voice Coding: FPGAs implement the ADPCM core and can process eight full duplex data streams and support the G.721, G.723, G.726, G.726a, G.727, and G.727a ITU standards.

## IP Phones

IP phones are telephones that connect to a LAN rather than the traditional phone jack. They are essentially a telephone with a built-in VoIP gateway and LAN interface circuitry. Processing functions in these systems are usually split between a DSP which handles voice processing functions, and a RISC processor which handles signaling, system management, and network protocol processing.

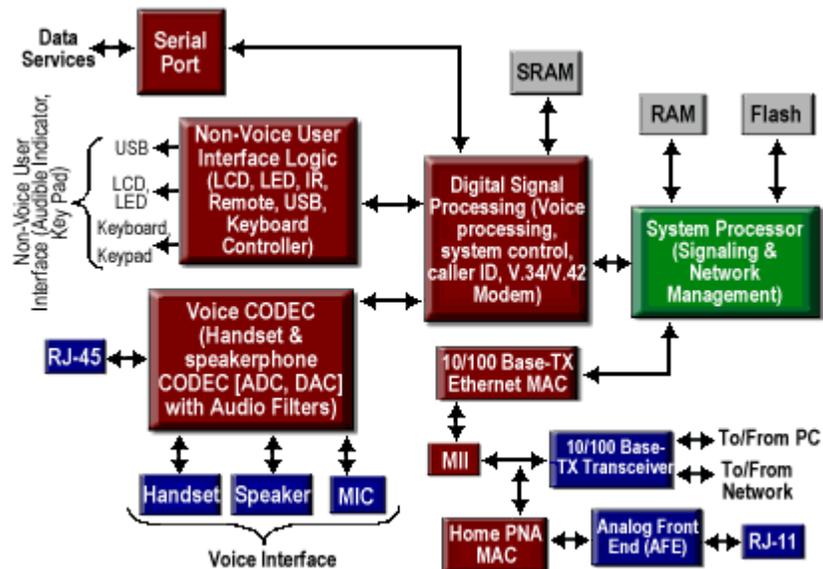


Figure 4: Phone Block Diagram

Figure 4 illustrates a typical IP phone architecture. It includes a voice codec (for analog to digital and digital to analog conversions), user interface logic (to interface to the keypad, status display, and audio indicator used for ringing) and the optional data (serial) port (for functions such as PDA synchronizing). Programmable logic solutions provide product differentiation and interface to multiple technologies, such as:

- ◆ System and user interface logic: PCI, RS-232 serial ports and other glue logic functions
- ◆ LAN, home networking, and wireless LAN interfaces: IEEE802.3, HomePNA, IEEE802.11, HiperLAN2, HomeRF
- ◆ DSP, voice codec

FPGAs implement the complex functions needed to interface network processors to switch fabrics or other ASSPs such as port interfaces in the infrastructure. FPGAs can be also be used as application specific coprocessors for network processors. In these applications, FPGAs are used to accelerate complex frame processing algorithms such as: traffic classification, traffic scheduling and shaping, complex policies, and queue management.

## Conclusion

Internet telephony has grown up, and is now part of the mainstream communication scene. While a complicated technology, it provides cost and bandwidth savings to the consumer and the enterprise. FPGAs present a low risk, cost effective way for system designers to develop and build VoIP gateways and solutions. Hence, FPGAs are enabling the convergence of data and voice.

## About the Author

Amit Dhir: System Architect, Strategic Applications, Xilinx San Jose



Amit Dhir's primary responsibilities as a Manager in the Strategic Solutions Marketing group at Xilinx include technical and market research, and evangelizing new and emerging markets. He is the author of "The Home Networking Revolution - A Designer's Guide," a book focussed on home networking technologies and components of the consumer and broadband access market. He has published several articles and white papers on topics covering the role for FPGAs in Wireless, Embedded, Telecom, Networking, and Consumer applications, and has presented at several industry conferences and consortiums.

Amit has a BSEE from Purdue University and a MSEE from San Jose State University.

### Articles by Amit Dhir

[FPGAs Driving Voice-Data Convergence 06/04/2001](#)

---

---

## Revision History

The following table shows the revision history for this document.

Date	Version	Revision
06/04/01	1.0	Initial Xilinx release.
11/21/07	1.0.1	Converted format from HTML to PDF.