MPEG-4 and the Future of Mobile Video

Software Development Forum Multimodal SIG

Robert Bleidt

March 4, 2004

streamcrest SM

Good evening, I'm Robert Bleidt with Streamcrest Associates, a new media consulting firm here in the Valley.

Tonight, I'm going to "set the stage" for our two primary speakers by explaining some of the history of mobile video and the technology behind it.

Agenda

1 | **What's Mobile Video?**
- The Early Years
- Current Innovation
- Future Alternatives

2 | **Video Codecs and MPEG-4**
- Why Compression?
- Codec Evolution
- MPEG-4: All Things to All People
- Scalability
- The Other MPEG-4
- The AVC Codec
- Royalties

3 | **Applications and Deployment Dilemmas**
- Bandwidth Costs
- Content Acquisition
- Operators vs. Services
- Cultural Acceptance

4 | **Introducing the Experts**
- Packet Video
- Mindego

First, we'll look at mobile video history. Then I'll explain how video codecs like MPEG-4 work and why we need them. Finally, I'm going to talk about some of the business issues we face in deploying mobile video.
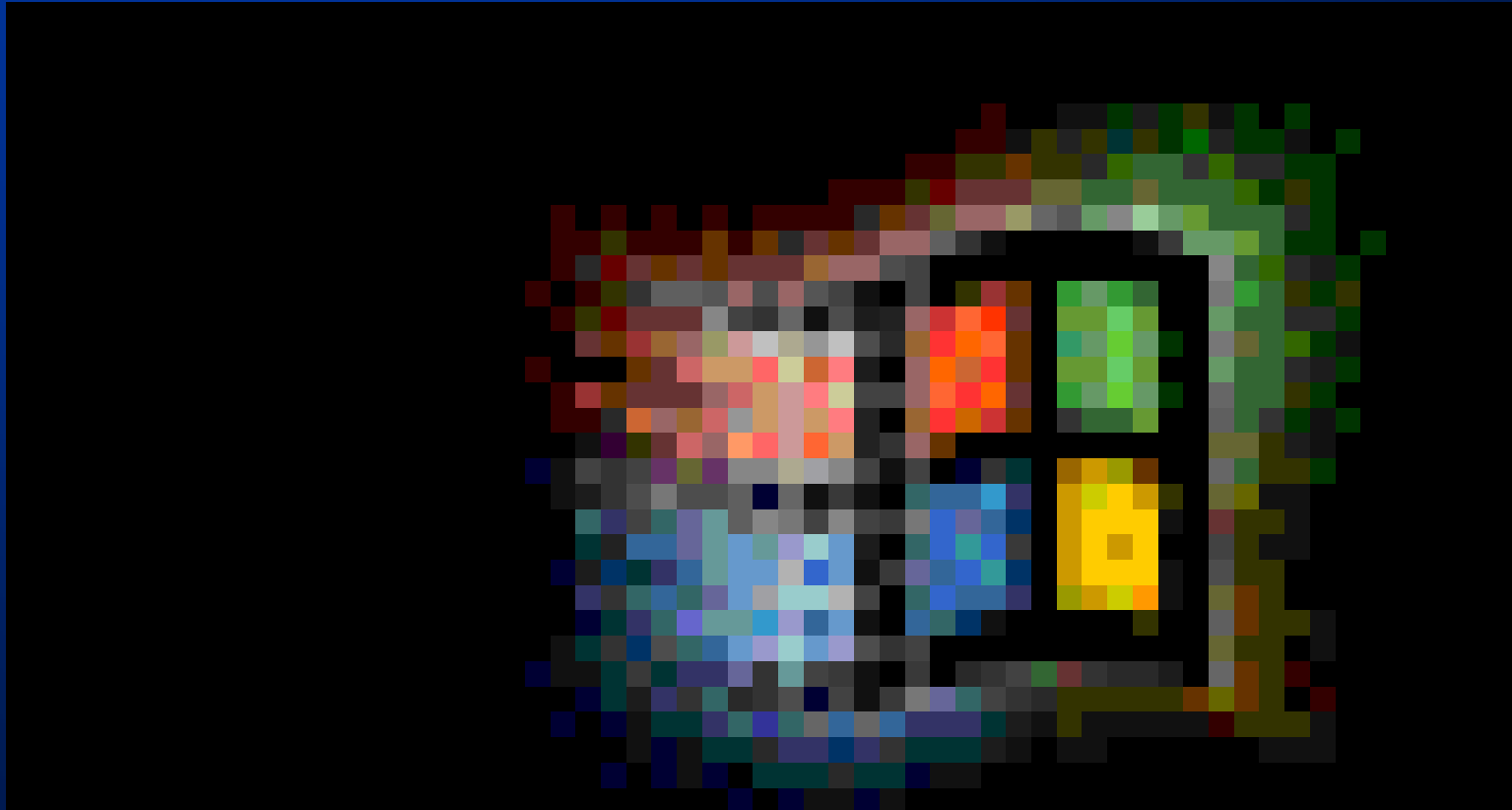
Mobile Video – The Early Years

streamcrest SM

Let me begin by explaining how mobile video started...

Dick Tracy's Wrist TV (1964)

http://www.comicspage.com/dicktracy/first-wristtv2.html

3/5/2004

streamcrest℠

Forty years ago, the creators of the Dick Tracy comic strip wrote the first product spec for a mobile video device, though it was not too practical at the time, needing a nuclear power source to operate. (but I'll bet the battery life was great...)

Sony's First Portable TV (1965)

- Sony 4-203w Battery-Powered Solid-State TV
  (apparently still working in 2000)
  3.5 inch CRT screen

www.tvhistory.tv

streamcrestSM

In the real world at that time, the best we could do was a lunch-box size TV with a handle.

Seiko TV Watch (1982)

- 1.1 inch LCD screen

- Display only, electronics in waist pack

- Obscure collector's item

www.cybercurator.com

streamcrest℠

Twenty years later, we had the display problem solved, but the electronics still required a belt pack to hold them.

Panasonic DVD-L10 DVD Player (1998)

- First portable DVD player

- 2 lbs, 6" square

- Initial price ~$1300

- About 5M portable units sold in 2003 (In-Stat)

Panasonic

Though products like Watchmans and other handheld TV's were brief fads, it was only about six years ago that we saw the first successful mobile video device, the portable DVD player.

The Early Years – some thoughts

- We are still waiting for Dick Tracy's goal of strap-on two-way videophony (… But stay tuned…)

- Dick Tracy was an enterprise customer with a need

- Consumers have accepted some mobile video concepts – like portable TVs and DVD players – in limited numbers

The portable DVD player proved that consumers are willing to pay for mobile video when they can experience a benefit. But it's a one-way stored-program device. What about Dick Tracy's dream of two way videophony?

Mobile Video – Current Innovation

The DoCoMo Story

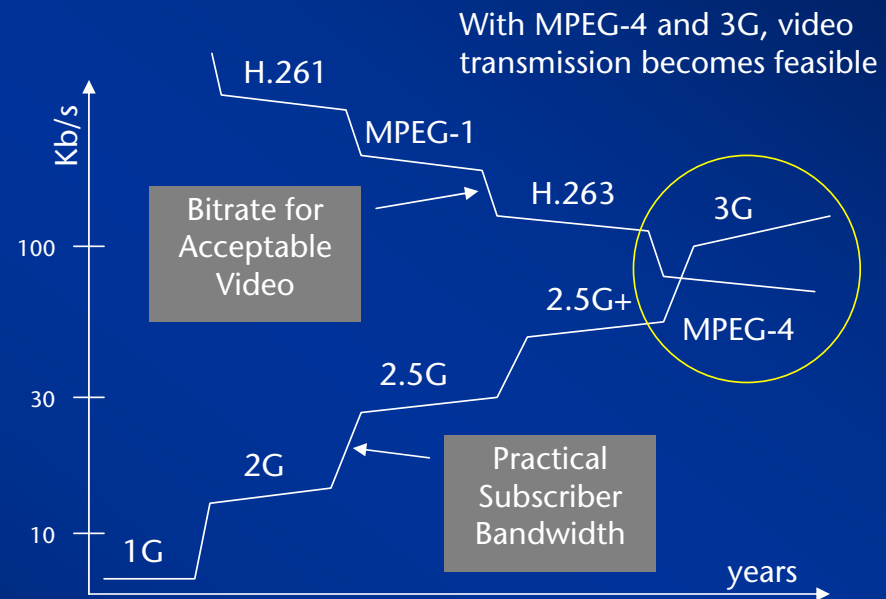## Hutchinson and Other Operators

streamcrest℠

As you'll see in a minute, Japanese consumers have been able to use two-way video phones for about three years, while other mobile operators are starting to deploy video phones in Asia and Europe.

Evolution of Codecs and Networks Enables Video

- The bitrate needed for acceptable video quality at handset resolutions has gone down as video codecs have improved...

- while the practical ($) bandwidth available to a subscriber has increased with new generations of mobile service

With MPEG-4 and 3G, video transmission becomes feasible

Kb/s

H.261

MPEG-1

H.263

3G

Bitrate for Acceptable Video

2.5G+

MPEG-4

100

2.5G

30

2G

Practical Subscriber Bandwidth

10

1G

years

Bitrates and times are approximate examples only, dependent on network load, subjective quality and other factors

This has become possible because progress in video compression technology and in mobile networks has finally made it practical to send acceptable video to a subscriber.

# Wireless Operators need Applications for revenue growth

- Wireless operators have saturated most markets with basic mobile phone service and have been looking to new service offerings to increase their average revenue per subscriber

- Handset manufacturers are currently stuck in a replacement market and would like to sell new products

- Thus – wireless operators purchased expensive 3G spectrum licenses to offer new services:

  - Web browsing

  - Still picture delivery (Multi-media Messaging Service)

  - Music Clip Download

  - Downloadable and Multiplayer Games

  - Videophony

  - Video Clips on Demand

So the technology has made it possible, but what about the business side? Mobile operators have saturated their markets with basic services and want continue growth by increasing their ARPU. Handset manufacturers are looking to offer new high-margin products instead of replacement handsets. So, the industry has attempted to introduce new services, with varying degrees of success. Today, everyone has bought a new phone with web browsing and MMS capabilities, but what about video? For that we have to look to Japan...

Japanese Mobile Market Leads Innovation

- Three main operators:
  - DoCoMo (NTT – Japan's original RBOC/PTT)
  - J-Phone (acquired by Vodaphone)
  - Au (KDDI – Japan's original Long Distance Carrier)

- First:
  - Web Browser
  - Java Applets
  - 3G Service
  - Videophony

Japanese Mobile Market Share (Rough Estimate)


- J-Phone 20%
- KDDI 22%
- DoCoMo 59%
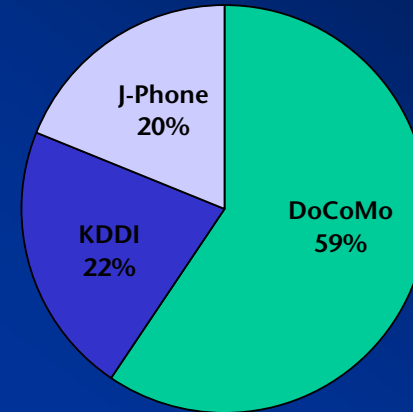
Average life of Japanese handset: six months

The Japanese market leads the world in development of many new mobile services, with the first practical mobile web browser, 3G service, and videophony. Most of these innovations have come from Japan's dominant mobile operator, NTT DoCoMo.

NTT DoCoMo Pioneered New Services

- iMode – Web Browsing integrated into phones with high resolution 65K-color displays
  - Wildly successful in Japan
  - Docomo purchased share in KPN (Netherlands) and AT&T Wireless (U.S.) to rollout service to other markets
  - Currently 40M subscribers, iMode ARPU $20/mo. on 10 web pages/day

iMode Subscribers

(Y-axis: Millions, 0 to 45; X-axis: Feb-99, Feb-00, Feb-01, Feb-02, Feb-03)
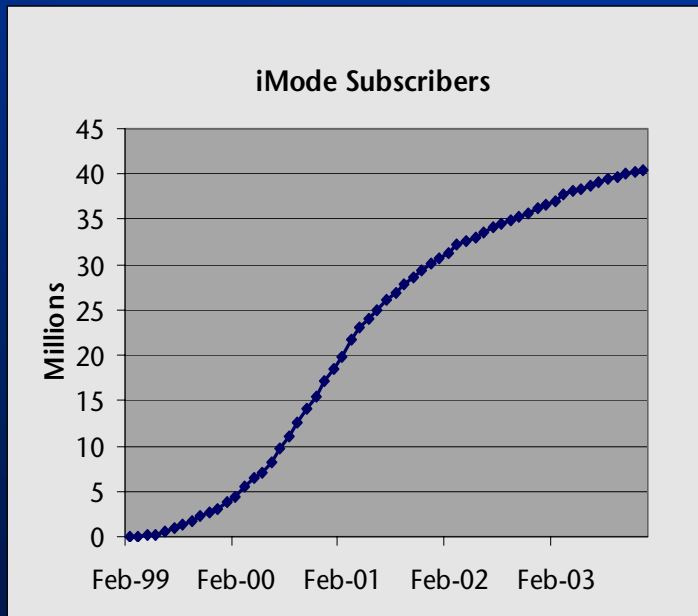
DoCoMo has suppliers private-label their handsets with the DoCoMo brand. D, F, SO, N indicate different manufacturers.

streamcrest℠

DoCoMo's first product innovation was a web browser combined with a handset that offered a high-resolution display and a walled garden of third-party web sites. Skeptics familiar with the WAP browser services being offered in Europe and America said DoCoMo would never get anywhere with it. They were still saying it a year later when DoCoMo had five million customers paying $10 a month. Today a third of Japan's total population are iMode subscribers.

DoCoMo's follow-on trial products were market failures/flops

- Eggy – Video mail and music download device

    – Oriented to youth market?

    – Product a flop

    – Uses ISO MPEG-4 video and GSM-AMR or ADPCM standard audio

    – Technical development with Microsoft

- Picwalk Portable Music Download Device

    – Also a flop

- Success with Java-based applets downloaded under iMode service – iAppli

streamcrest℠

When DoCoMo tried to move beyond web browsing to music and video, their initial attempts were failures – they hadn't found the right product for the youth market they focused on. Analysts were calling them a one-trick pony, forgetting that all that iMode revenue was going to let them make a big bet...

In Nov. 2001, DoCoMo Pioneered Videophony with MPEG-4 on their 3G FOMA Service

- This was a big bet on the utility of high-speed packet radio networks for mobile web and video.

- Each handset ($300-500) had a video camera and large display

- MPEG-4 video is used at 64 Kb/s

- Service area is now 99% of population.

FOMA N2002 Handset

Compatible with i-mode services, including i-motion, i appli, i-mode mail and various online services.

2.2-inch full-color (65,536 colors) liquid-crystal display

Continuous stand-by approx. 55 hours   Continuous talk approx. 90 minutes

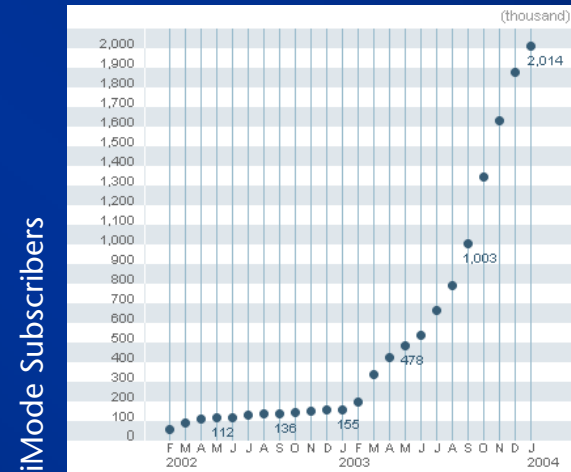Packet transmission speed Downlink max. 384 kbps / Uplink max. 64 kbps

Circuit-switched transmission speed 64 kbps

streamcrest℠

In November 2001, DoCoMo introduced their FOMA service, which offered full two-way video communication over a new true 3G network, at a time when European carriers were saying 3G service was years away. Initially, the battery life of these phones was only about an hour, and the 3G network only covered central Tokyo. Given a subscriber cost of about $500 per handset, in the first year DoCoMo only signed up 50 or 100 thousand subscribers. The skeptics were sure DoCoMo had another flop on their hands.

FOMA Becoming Another Hit for DoCoMo

- Initial sales were slow, recent hockey-stick to about 2M subscribers

- Third series of handsets introduced in February 2004: 80% of initial weight, 3x battery life

- 320 x 240 QVGA display, fingerprint sensor, TV output

- 2M pixel camera, Office document viewer on some models

- Flash web browser (iMode) and 500KB Java (iAppli) capacity

(thousand)

iMode Subscribers

2,014

1,800

1,003

478

112   136   155

F M A M J J A S O N D J F M A M J J A S O N D J
2002   2003   2004

Review Draft
Confidential Client Presentation © 2004 Streamcrest Associates

3/5/2004

streamcrest℠

Well, in the past year or two, the FOMA service has grown to two million subscribers on a network that covers 99% of the population. Last month, they introduced their third generation of handsets, with battery life comparable to voice units, and a 320 x 240 pixel display.

Hutchinson-Whampoa

- Sold Orange network to Mannesmann nee Vodafone in 1999, has no 2 G network except HK – a "greenfield" operator

- JV with DoCoMo on HK and UK 3G

- 3G service in Hong Kong, UK, Australia, Austria, Italy, Sweden

- R&D partnership with DoCoMo

- UK "3" service started March 2003:
  - Video call: £050/minute
  - Premier League halftime highlights: £0.70
  - Top Shelf one-minute video clip: £1.50
  - Voice Call: £0.05-0.25
  - Voice Plans: 500 min/ £25, 750 min/ £35
  - 37,000 subscribers through June 2003
 - Content: Reuters, ITN, Playboy, BBC Technology

Handsets supplied by NEC and Motorola, NEC order increased to 2M units, Sony may be future supplier

streamcrest℠

About a year ago, Europe saw it's first large deployment of 3G service by Huchinson, which includes videophone handsets.

Rest of World Tire-Kicking Multimedia

- Many wireless operators conducting trials of multimedia delivery using GPRS or CDMA 1X technology
  - Verizon
  - Sprint
  - AT&T
  - Deutsche Telekom
  - Vodaphone
  - KPN

- Test platform is universally iPaq Pocket PC, usually with CF card or PCMICA modem
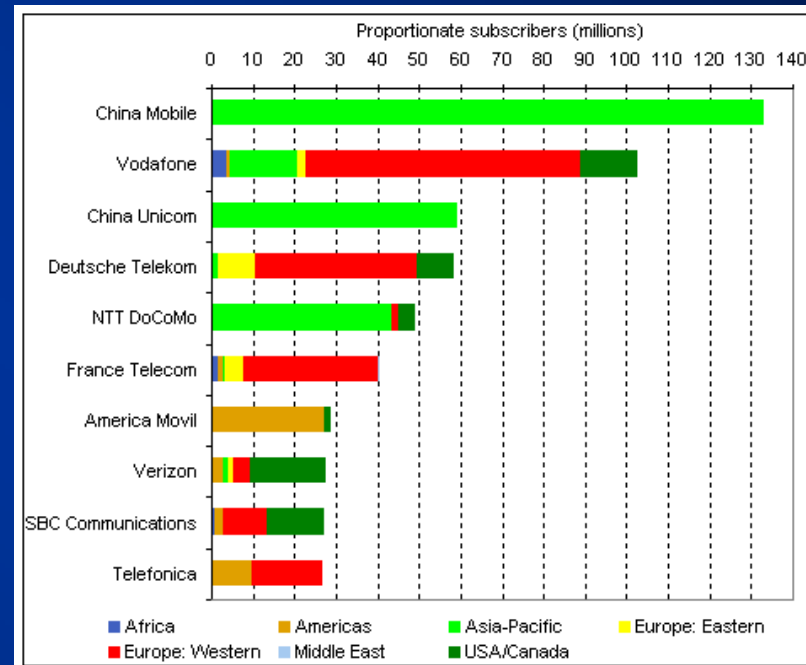
EMC World Cellular Investors, Dec. 2002

While innovators DoCoMo and Hutchinson have successfully deployed videophony, the real market will depend on what Vodafone and the other top-ten carriers decide to do.

So what about the world's other mobile operators? Operators in Hong Kong and Korea have closely followed DoCoMo and Japan's two other providers, but the rest of the world is still in the tire-kicking stage, conducting limited trials to PDAs.

Rollout of 3G services hits debt and doom

- Aside from DoCoMo and Hutchinson, most operators:

  – Have no vision of or experience with new services – they are used to selling airtime and handsets, not services

  – Are mired in problems of voice service deployment using cheaper 2.5G technologies

  – Are giving lip service to new services with trials

  – Synergistic dreams of content/operator convergence (ex. Vivendi) have dissipated with the .com crash

  – Will not deploy video services until their debt loads and voice deployment problems improve and they regain financial favor

  – Long-term

    - Operators own the customer

    - Mobile is so mired in balkanized standards that operators can maintain a walled garden of services (DoCoMo looking to be a winner as aggregator)

    - Surviving operators will offer services based on consumer acceptance and demand

    - APAC deployment will lead EU, US delayed due to frequency allocations

Why haven't they kept up? One important reason is that, unlike DoCoMo, they have no experience with developing or deploying real services instead of ring tones and airtime. Thus, they don't have people who can envision new services and make them happen.

Many of these operators still have debt and implementation problems deploying their 2.5 G networks. And their dreams of developing services by big mergers with content companies have unraveled with the dot-bomb market. Long-term, though these operators will eventually follow.

Microsoft Smartphone is shunned

- Microsoft needs to enter mobile or CE market to continue growth

- Despite PDA – Phone convergence, operators are staying away from Microsoft's Smartphone initiative

- 2Q03 EMEA smartphone shipments: 94% Symbian, 6% MS (canalys/WSJ)

- Orange – Current partner for Smartphone SPV from HTC – "not in our interest to have any one dominant operating system" (also announced deal with Palm)

- Deutsche Telekom – Dropped plans to deploy SPV in May 2003

- Sendo – European private-label handset manufacturer, recently withdrew from Smartphone SPV launch, said "Microsoft impossible to work with"

- AT&T Wireless may be next partner

- IDC Analyst: market expects carrier-grade reliability (which Smartphone is perceived not to provide)

streamcrest℠

Another company that can produce innovative products when it wants to is Microsoft. Microsoft sees digital media as a key technology to enter new markets, and mobile handsets is one they would like to conquer. Unfortunately, the mobile industry is very cautious about working with Microsoft, and Microsoft's products so far have not met their reliability expectations.

Feeble U.S. Attempts Gain Mixed Reviews

- Sprint MobiTV/Idetic
  - "we're completely underwhelmed by the prospect of paying ten bucks a month to watch choppy, one frame per second video on a cellphone" – Gadget Site Gizmodo.com
  - "this hot little application, even with its limited frame rate and channel selection, might turn out to be a bigger success than the other applications combined" – PC Magazine
  - "a pitiful excuse for live TV on a dinky screen at a frame rate of one (1) frame per second" - Fortune

Meanwhile, here in the U.S., we don't have the 3G networks needed for video, and our operators have been experimenting with "toy technology" that can send one frame a second over existing networks.

Mobile Content Convergence Devices

This is Streamcrest's coined phrase for new portable media capture and consumption appliances now being developed

streamcrest℠

Remember our look at the DVD player? Mobile video is not just telephone handsets – another type of consumer mobile video device is what we've termed Mobile Content Convergence Devices, or MCCD's.

What's a MCCD?

- New class of devices incorporating multiple functions or new storage media to converge different applications into portable device

- Include video capture or playback

- Leisure or capture-oriented instead of data or communications-centric (as in smart phone devices)

- Typical features:
  - Camcorder
  - Still Camera
  - MP3 Player
  - Dictation Recorder

An MCCD is more leisure or capture-oriented than a phone handset, and typically includes the features of a camcorder, a still camera, and a music player in a compact, mobile device. Often a MCCD will use flash memory or a hard disk drive instead of traditional tape or optical storage.

Example MCCD Products (2003)

Recently Introduced Mobile Content Convergence Devices

**Panasonic SV-AV30U
e-ware Convergence Device**

**Samsung ITCAM-7
Convergence Device**

	Panasonic SV-AV30U	Samsung ITCAM-7
Video Recording and Playback	Fixed focal lens, ¼" CMOS sensor, 2" 200,000 pixel display	10X optical zoom lens, 350,000 pixel CCD, 2" 210,000 pixel display
Video Format	MPEG-4	MPEG-4
Video Resolution	320x240 (SIF), 15 fps	640x480, 30 fps
Video Bitrate	Nominal: 250 kb/s @ QSIF 1 Mb/s, 400 Kb/s @ SIF	3.0 Mb/s at highest res. (other rates unknown)
Music Format	MP3, AAC (no encoding)	MP3
Voice Format	G.726	Stereo, unknown
Still Capture	JPEG, 640x480	JPEG, 640x480
Storage Media	SD Card	1.5 GB Hard Disk, memory stick
Other Features	Docking cradle for connection to A/V system, timed recording	"data tank", "PC camera", USB 2.0
MSRP	$399 (model less remote and cradle, $299)	$649
Author has seen:	Working Product, NAB	Static Model, CES

streamcrest℠

The first of these products appeared last year. One of the most popular has been the Panasonic "e-Wear" series that uses MPEG-4 to record onto flash memory.

Additional MCCD Products

Archos AV340 Convergence Device

Intel Personal Video Player Reference Design

streamcrest℠

Some MCCDs are closer to music player devices, and may have capture functions in an external adapter.

CES 2004 Introductions

Creative Labs PVP

Audiovox PVP Line

Alaska Airlines PVP (at TI booth)

streamcrest℠

In January, Microsoft introduced its own MCCD, it Personal Media Player platform, though manufacturing partners such as Creative Labs. This device is playback-only, and is tightly coupled to a Windows XP PC.

Mobile Video – Future Alternatives

DVB-H and DMB

streamcrest℠

So a Panasonic MCCD will let us carry our pre-recorded programs with us, and we can use a DoCoMo handset to make a video call. But what if we want to watch a long-form live TV program?

Broadcast popular content instead

Sanyo ISDB Handset (prototype)

NEC ISDB Handset (prototype)

- Popular live content doesn't require a point-to-point connection – it could be broadcast over a common channel

- Several initiatives underway to do this:

  - DVB-H (formerly DVB-X) – use part of European DVB digital TV signal to carry stream to handset. (Nokia)

  - 3GPP MBMS (Multimedia Broadcast Multicast Service) – use a dedicated channel on 3G network to broadcast stream to handset (Vodaphone)

  - DMB (Digital Multimedia Broadcasting) - (Samsung)

  - Japanese Handsets with ISDB (DVB variant) digital TV receivers (Sanyo, NEC)

streamcrest℠

There are several initiatives in development to use broadcast networks to send TV content to mobile handsets. These range from simply a in-handset TV tuner for standard digital TV broadcasts, to ones where a portion of a terrestrial or satellite data stream is used to send a useable signal to a handset on the move.

Agenda

1	What's Mobile Video?	• The Early Years • Current Innovation	• Future Alternatives
2	Video Codecs and MPEG-4	• Why Compression? • Codec Evolution • MPEG-4: All Things to All People	• Scalability • The Other MPEG-4 • The AVC Codec • Royalties
3	Applications and Deployment Dilemmas	• Bandwidth Costs • Content Acquisition	• Operators vs. Services • Cultural Acceptance
4	Introducing the Experts	• Packet Video • Mindego	

streamcrest SM

Why do we need video compression?

And how does it work?

streamcrest℠

Earlier I mentioned that video bitrates have finally shrunk to the level that new mobile networks can support. So, how does this work?

Compression is the Instant Coffee of Media

Original Content

Encoding → Decoding →

Compressed File or Bitstream

Reconstituted Content

- Encoders "freeze dry" the signal by removing information you won't notice and cleverly packing the rest

- The resulting compressed file is easy to store and transmit

- Decoders "add water" by undoing the packed format and reconstructing the picture

- Just like instant coffee, it's close enough to the real thing some of the time

With the lossy compression we use in video, the idea is to remove information you won't usually notice and do a very clever job of packing the rest into an encoded bitstream, which is then reconstructed at the receiving end.

Is Compression Necessary?

Height

Width

Frame Rate

With uncompressed video, each pixel in every frame is transmitted unchanged. The bitrates needed are fine for processing in equipment, but too high for practical storage or transmission.

Uncompressed Studio Video Example using American Standard-Definition TV:

Height:		483 lines
Width:	x	720 pixels
Frame Rate:	x	30 frames/sec
Color Depth:	x	20 bits/pixel
Framing Overhead:	x	1.3
Total Bit Rate	=	270 Mb/sec

From SMPTE 259M Standard

streamcrest℠

Keeping in mind that we may have 50-100 Kb/s available for transmitting a video clip, the 270 Mb/s of a standard-definition TV signal clearly needs compression. So how do we do it?

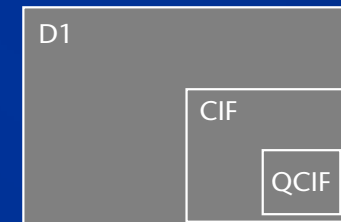# Easy Compression #1 – Reduce Image Size

If the bitrate is too high, the simplest way to reduce it is to throw away part of the picture.

Common sizes for mobile video are ½ to 1/5 standard TV resolution.

Resolution	525-line Video "NTSC" – U.S.	625-line Video "PAL" - EU
SQCIF	128 x 96	128 x 96
QCIF	176 x 144	176 x 144
CIF	352 x 288	352 x 288
½ D1	360 x 486	360 x 576
¾ D1	540 x 486	540 x 576
VGA	640 x 480	640 x 480
D1	720 x 486	720 x 576 (4CIF)

~~Uncompressed Studio Video Example using American Standard-Definition TV~~:

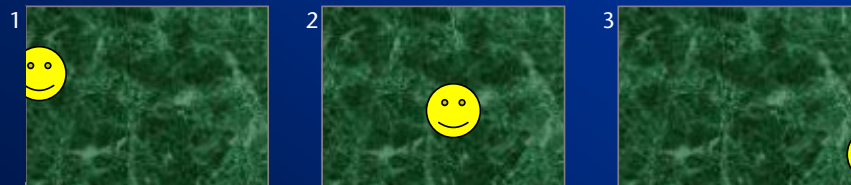|		
Height:		144 lines
Width:	x	176 pixels
Frame Rate:	x	30 frames/sec
Color Depth:	x	20 bits/pixel
Framing Overhead:	x	1.3
Total Bit Rate	=	20 Mb/sec

D1

CIF

QCIF

First we can start with some easy things – like reducing the size of the picture. Most mobile video uses the QCIF format that's about a sixteenth of the pixels in a studio TV signal. This can get us to 20 Mb/s.

Easy Compression #2 – Reduce Frame Rate

Reducing the frame rate has no effect on images of stationary objects, but motion of objects is jerky (judder)

~~Uncompressed Studio Video Example using American Standard-Definition TV~~:

Height:		144 lines
Width:	x	176 pixels
Frame Rate:	x	15 frames/sec
Color Depth:	x	20 bits/pixel
Framing Overhead:	x	1.3
Total Bit Rate	=	10 Mb/sec

1   2

Instead of:

1   2   3

streamcrest℠

Another easy thing is to reduce the frame rate, which doesn't reduce the resolution like cutting the image size, but will make motion appear jerky.

Easy Compression #3 – Reduce Colors

Normal consumer TV uses 8 bits or 256 possible values of red, green, and blue color components to represent a pixel, or 24 bits per pixel. (This studio example has used 10 bits so far.)

Because color (as opposed to monochrome) information is shared between two pixels, due to chrominance subsampling, only 16 bits are required.

Many LCD displays are incapable of displaying a full 256 steps of each color.

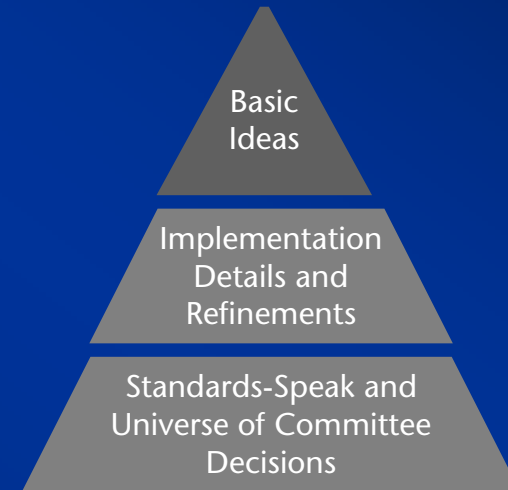| ~~Uncompressed Studio Video Example using American Standard Definition TV:~~		
Height:		144 lines
Width:	x	176 pixels
Frame Rate:	x	15 frames/sec
Color Depth:	x	14 bits/pixel
Framing Overhead:	x	1.3
Total Bit Rate	=	8.6 Mb/sec

8 bits/color

7 bits/color

6 bits/color

5 bits/color

Another easy step is to reduce the number of bits in each pixel. (show depth of projector at site) This gets us to 8.6 Mb/s, a long way from 100 Kb/s – so...

Easy tricks aren't enough – applying some science:

- Data Compression Techniques
  - PK Zip, Huffman Coding
  - Adaptive Differential PCM

- Frequency Domain Transforms
  - JPEG
  - Motion JPEG / DV / Digital Betacam

- Motion-Compensated Transforms
  - H.261, H.263
  - MPEG-1,-2,-4,AVC

Basic Ideas

Implementation Details and Refinements

Standards-Speak and Universe of Committee Decisions

We have to apply some heavy science to the problem. Instead of getting stuck in the details, I'm going to present just the basic ideas. Understand these three ideas and you'll know how MPEG works...

Science Trick #1 – Data Compression

Original Data:

00 00 00 00 11 23 23

Run Length Coding:

04 00 01 11 02 23

Repeat Count
(Run Length)   Value

Typical Codes, English Letters:

	ASCII	Huffman
E	01000101	100
A	01000001	0100
Z	01011010	1110111111

- Instead of sending our data, we send a coded version that is shorter

- With Run Length Coding, we replace a repeated "run" of a value with a count and the value

- Huffman coding replaces a value with a variable-length code based on statistics. Popular values get a shorter code

streamcrest SM

Our first idea is to use techniques from information theory to reduce the number of bits we send. Instead of a sending a bunch of identical values, we send just a repeat count. And instead of using a simple code with the same number of bits for each value, we use variable-length codes with shorter numbers of bits for the most popular values.

Science Trick #1 – Application: PNG, GIF
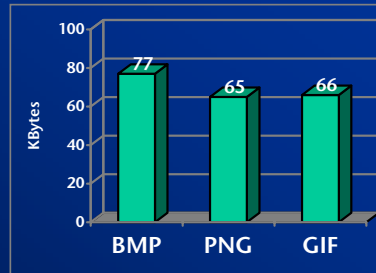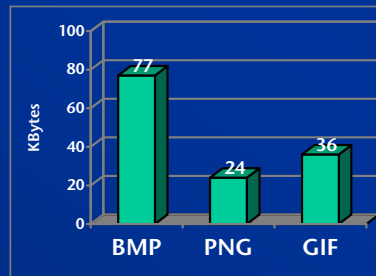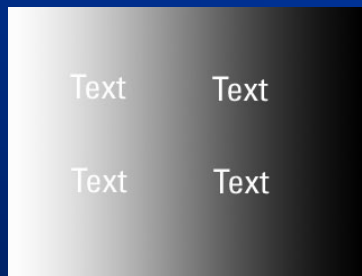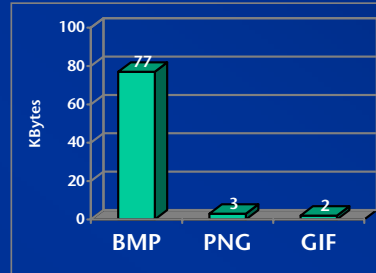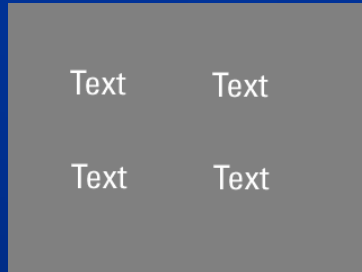
- Simple images with many repeated pixels can be compressed 20-40x in PNG or GIF formats

- Photo or video images can't be compressed much since they have few repeating patterns – a GIF of random noise is larger than the original

These are the ideas behind GIF and PNG images that you see on web pages. Simple images, like graphics with many repeated pixels, can be compressed twenty or forty times. But this doesn't work well as the image becomes more random, like a video image.

Science Trick #2 – Frequency Domain

Each pixel represents a frequency band, just like those on a CD player display

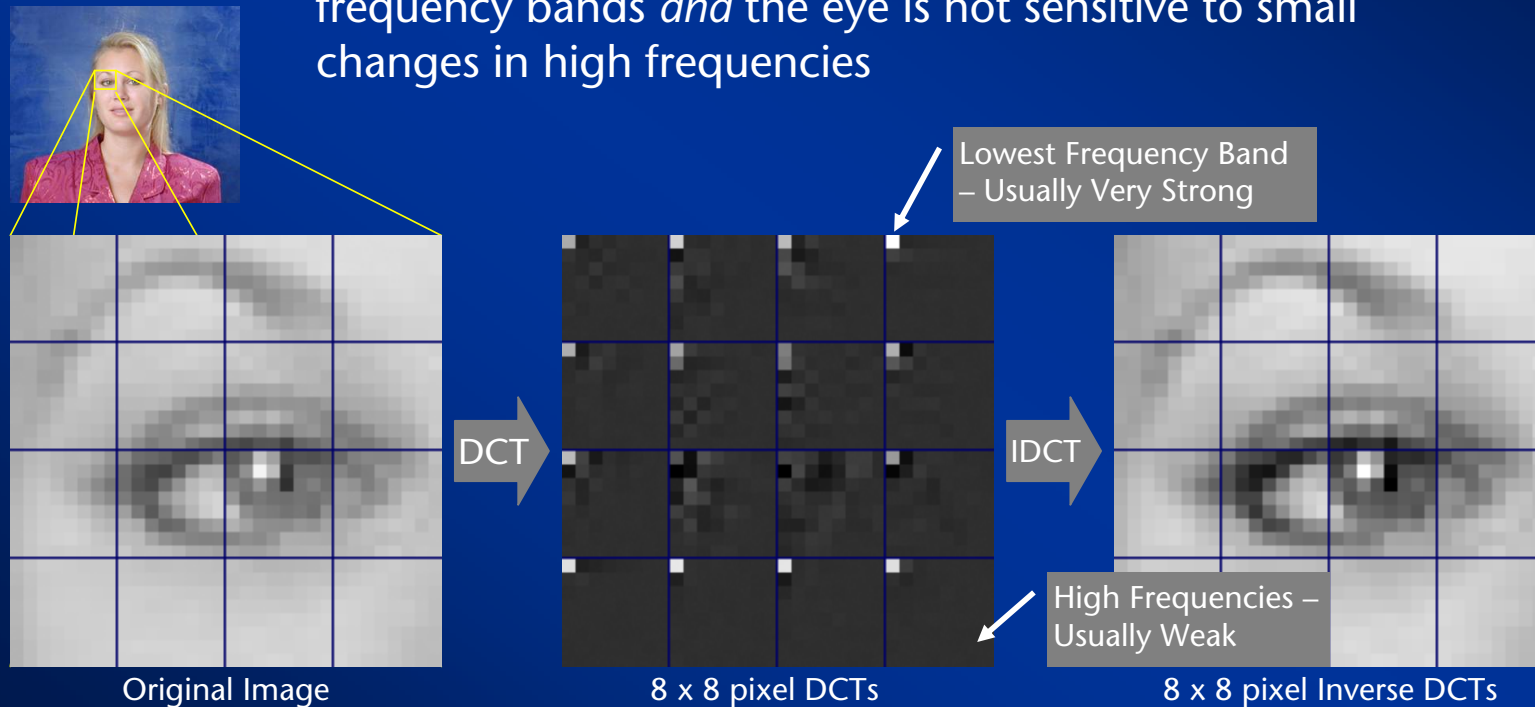| 8 x 8 Pixel Piece of Image | DCT Frequency Transform | 8 x 8 Pixel DCT of Piece | Inverse DCT Frequency Transform | 8 x 8 Pixel Piece of Image |
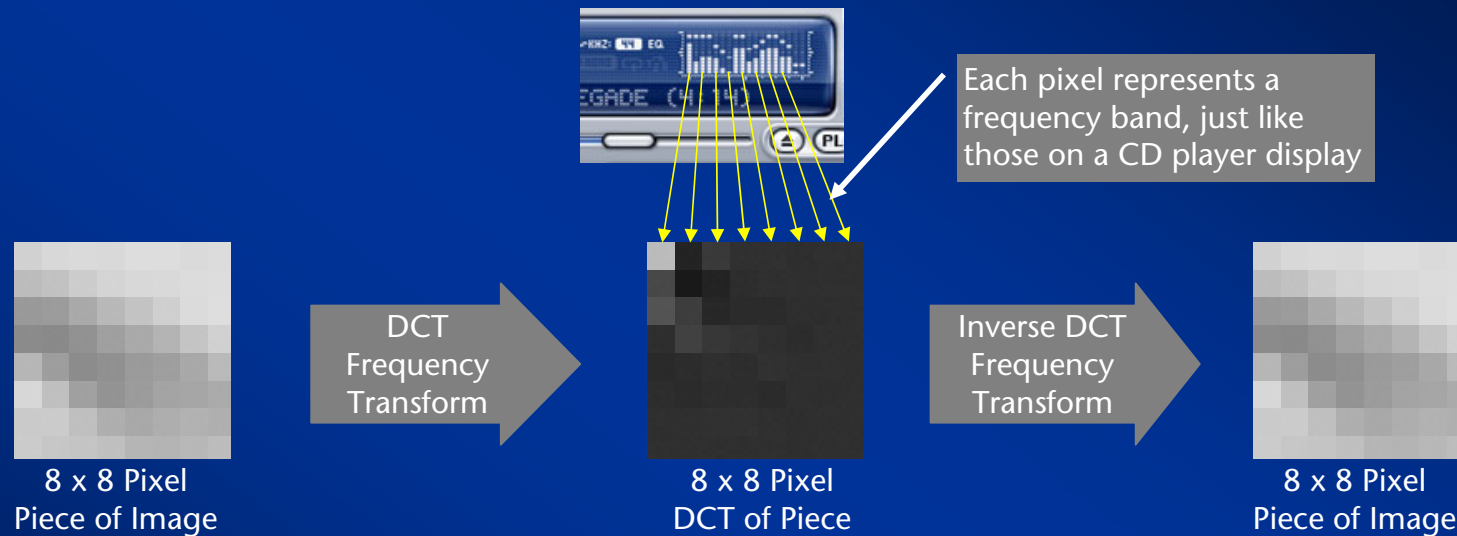
- A mathematical transform, or formula, lets us convert a block of pixels into another block where pixels represent the strength of frequency bands in the image.

- An inverse transform on this new block exactly re-creates the original image.

To get around that problem, we use another trick – transforming the image into the frequency domain. While this may sound complicated, it's really just a mathematical operation that lets us convert a block of pixels into another block where the pixels represent frequencies in the image, just like the equalizer bars on a stereo display. Since images are two-dimensional we end up with a block of pixels instead of a bar graph.

Science Trick #2 – Frequency Domain

- There is no magic in using the frequency domain – it's just that *most* images have little amplitude in the upper frequency bands *and* the eye is not sensitive to small changes in high frequencies

Lowest Frequency Band – Usually Very Strong

DCT

IDCT

High Frequencies – Usually Weak

Original Image

8 x 8 pixel DCTs

8 x 8 pixel Inverse DCTs

Transforming an image to the frequency domain is useful only because most of the high-frequency pixels in the transformed image are close to zero in most cases. And, the eye won't usually notice if we use a smaller number of bits per pixel to represent them.

# Science Trick #2 – Application: JPEG

Uncompressed BMP

77 KB (8.1 bpp)

DCT (#2)

Data Compression (#1)

JPEG from Photoshop Q=9

18 KB (1.9 bpp)

Errors multiplied by 16
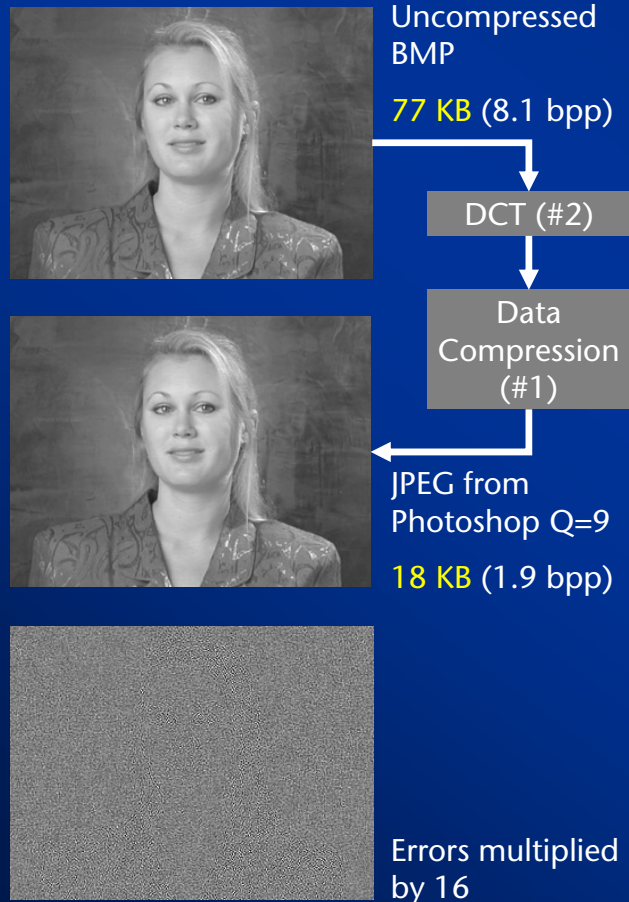
- The frequency transform doesn't reduce the number of pixels, but it does redistribute their values so that high frequencies have usually small values.

- We can use a lower number of bits to represent these high frequency pixels since they are small and the eye is not very sensitive to slight errors in their value. In some cases, we can just set them to zero.

- Data compression techniques can then compress these pixels very efficiently as there are then many repeated zero values and a smaller number of bits per pixel.

- Combining trick #1 and trick #2 – we have the basics of the JPEG standard.

streamcrest℠

Now we have an image with many repeated values that we can compress with our data compression tricks – but notice we've thrown some information away. Combining tricks one and two, we have the JPEG standard used in still cameras. In this example, applying these two tricks takes us from 8 bits per pixel to 2, but the picture appears just like the original.

Science Trick #3 – Frame Differences
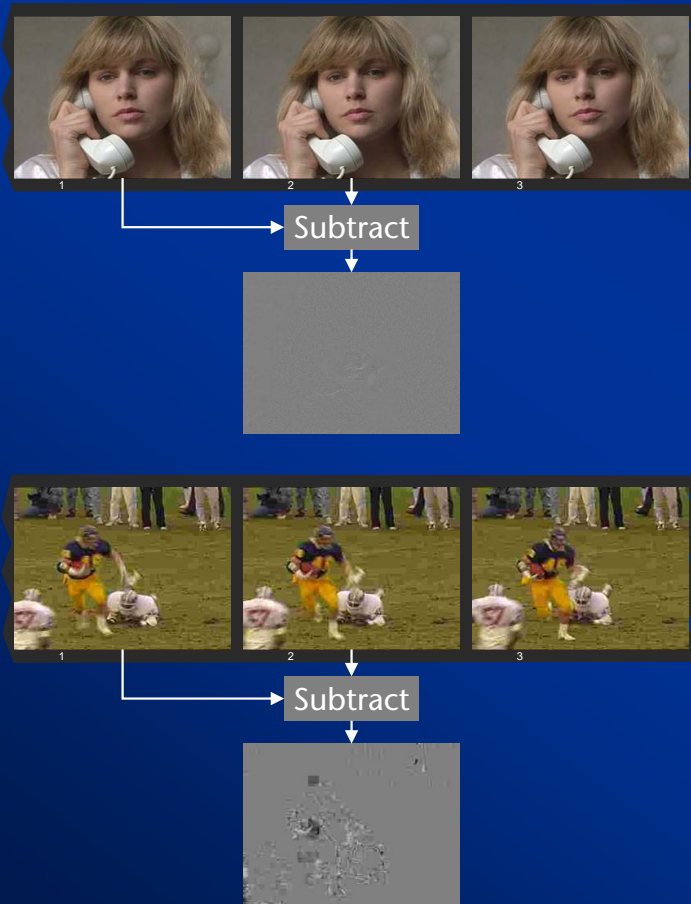
- Most videos have little motion from one frame to the next

- We can subtract the previous frame from a frame and encode only the differences, which are usually quite small

- As a refinement, we can move a piece of the previous frame a few pixels before subtraction to reduce the resulting difference image (The amount of movement is sent with the encoded image)

streamcrest℠

Trick number three is to take advantage of the fact that a video picture doesn't change much from one frame to the next. So, we can subtract the previous frame from current frame and encode only the differences. Even for a fast moving scene like the second example, most of the image is the same. (and if it's not – the motion will be so fast our eye can't tell)

Science Trick #3 – Frame Differences

Original Image (12.7KB JPEG, Q=5)

Previous Frame Subtracted from Image (6.3KB JPEG, Q=5)

- The difference frame usually has few details and is easy to encode with JPEG-like frequency domain encoding using tricks #1 and #2.

- If we just JPEG the difference frame using Photoshop, it is about ½ the size of a JPEG of the original.

- A combination of other minor tricks, plus a reduced "JPEG" quality setting makes frame difference coding have an efficiency of about 0.5 to 0.2 bits/pixel.

streamcrest℠

This trick allows us to cut our bits per pixel in half, and we can also cut the JPEG "quality setting" since the eye won't notice the loss of detail on a moving object. This gets us in the range of .5 to .2 bits per pixel.

Science Trick #3 – Application: MPEG

- The MPEG 1,2, and 4 standards use tricks #1,#2, and #3 combined to encode video

- Typical standard-definition video:
  - Motion-JPEG: 25 Mb/s (2.5 bpp)
  - MPEG-2: 5 Mb/s (0.5 bpp)

- MPEG-2 applications:
  - DVDs
  - Satellite and Digital Cable TV
  - HDTV
  - Tivo and PVRs

Approximately 50M Set-Top Boxes and 100M DVD players shipped annually worldwide using MPEG-2 encoding

These three tricks are the core of the MPEG 1, 2, and 4 standards. MPEG-2 is something we use everyday when we watch a DVD or digital TV.

Compression tricks put us in the ballpark

- With a good MPEG-1 or MPEG-4 encoder we can probably get tolerable quality at 0.2 bits per pixel.

- If we also reduce the frame rate to 7.5 fps, we can plan on a total bitrate of 45 Kb/s

- There is no fixed bitrate for a given situation, since

  - Encoders differ in quality

  - Some content is more difficult to encode cleanly

  - Desired quality levels are subjective and variable

~~Uncompressed Studio Video Example using American Standard Definition TV:~~

Height:		144 lines
Width:	x	176 pixels
Frame Rate:	x	7.5 frames/sec
Color Depth:	x	0.2 bits/pixel
Framing Overhead:	x	1.2
Total Bit Rate	=	45 Kb/sec

This example doesn't include audio, which can add 5-20Kb/s for speech or 32-128 Kb/s for music

Back to our example, if we assume we have a really good MPEG encoder with .2 bits per pixel, and we decide to cut the frame rate some more, we're in the ball park of the bitrates we can afford to send to a mobile handset.

Understanding Video Codec Evolution

streamcrest℠

So what's the difference between MPEG 1,2,and 4? And what about proprietary formats like Windows Media?

Timeline of Standardized Video Coding

	ITU –	ISO –	Proprietary Formats
	International Telecommunications Union	International Standardization Organization	
	Telephone, Radio, TV	Photography, Computer, Consumer Electronics	Successful standards have enabled an application – Video CD and early multimedia for MPEG-1, Digital TV for MPEG-2.
1984	H.120		
1990	H.261 – Video Conferencing		
1993		MPEG-1 – Video CD	
1994	(H.262)	MPEG-2 – Digital Cable and Satellite TV	
1995	H.263 – Improved Video Conferencing		Real Audio
1997		ATSC – U.S. HDTV	Real Video Windows Media
1999		MPEG-4	Unlike previous standards, significant proprietary formats deployed before MPEG-4.
2002	AVC (H.264)	AVC (MPEG-4 Part 10)	

Proprietary Standards have always developed faster due to lack of politics, bureaucracy, and review.

The tricks I've described have been used in video coding for about fifteen years, and the MPEG-1 and 2 standards were very successful because they enabled new applications. But standards bodies don't always move quickly, and five or six years ago, we saw proprietary codecs from Microsoft and Real Networks arrive before the MPEG-4 standard.
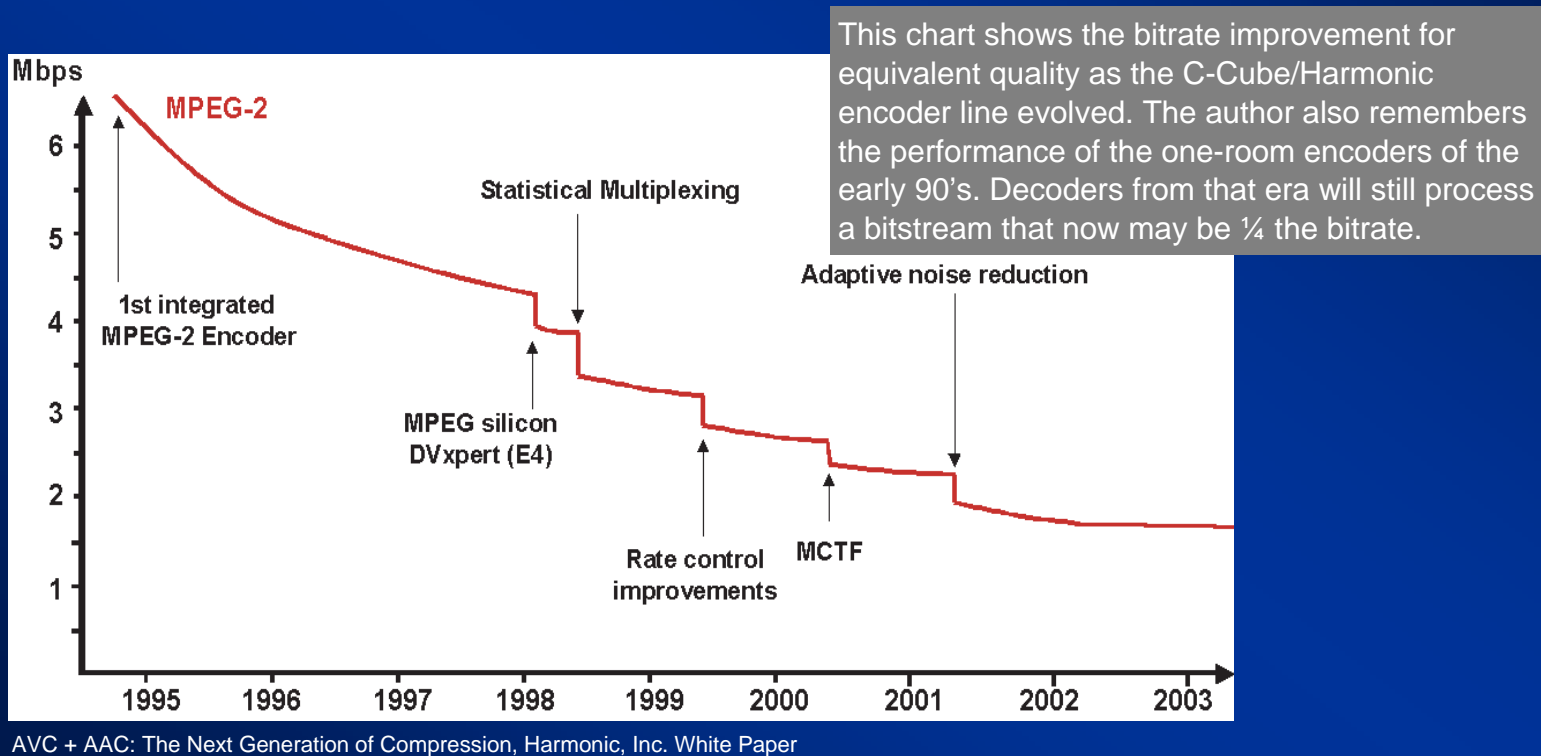
Proprietary codecs can innovate at will

- If there are no magic techniques, how can proprietary codecs be better?

    – They don't have to be compatible with existing decoders or interoperate with other vendor's products

    – There is no need for lengthy study, testing, and review by standards bodies

    – Political or IPR concerns are resolved privately

    – They can take advantage of features of specific platforms or be coupled with related technology, such as digital rights management systems or noise reduction techniques.

- Basically, they use the same techniques that standards-based codecs do, but with quick private implementation. Some, such as Windows Media and Real Video are very good products that seriously challenge MPEG-4

In some cases, these proprietary codecs offered better performance than the later MPEG-4 standard, because they can make changes faster and don't have to worry about compatibility. This turned out to be a shock for MPEG-4 licensors, as we'll see later.

Encoder innovation continues after standardization

- MPEG standards only specify what the decoder must do. The encoder may use any techniques to encode the content. This allows continued innovation and proprietary advantage while maintaining interoperability.

This chart shows the bitrate improvement for equivalent quality as the C-Cube/Harmonic encoder line evolved. The author also remembers the performance of the one-room encoders of the early 90's. Decoders from that era will still process a bitstream that now may be ¼ the bitrate.

Mbps

MPEG-2

Statistical Multiplexing

Adaptive noise reduction

1st integrated
MPEG-2 Encoder

MPEG silicon
DVxpert (E4)

Rate control
improvements

MCTF

6

5

4

3

2

1

1995   1996   1997   1998   1999   2000   2001   2002   2003

AVC + AAC: The Next Generation of Compression, Harmonic, Inc. White Paper

One thing to keep in mind, though, is that even open-standard codecs continue to get better over time. The standard only specifies what's in the decoder, and the encoder can be improved later, as we see here with the 3x improvement of MPEG-2.

MPEG-4 part of 3GPP industry standard

- MPEG-4 specified as one video codec for 3G mobile communications by 3GPP standards group

- Speech coding is almost exclusively GSM-AMR, not MPEG-4 CELP or AAC

- Telecom carriers like open, stable, interoperable, international standards – MPEG-4 appeals to them

- Handset suppliers have struck deals with Real and Microsoft, but are also considering MPEG-4

- Several chip houses building MPEG-4 codec/display driver chips for handset use

Today, carriers have experimented with Real and Windows Media, but most deployments have stuck with MPEG-4.
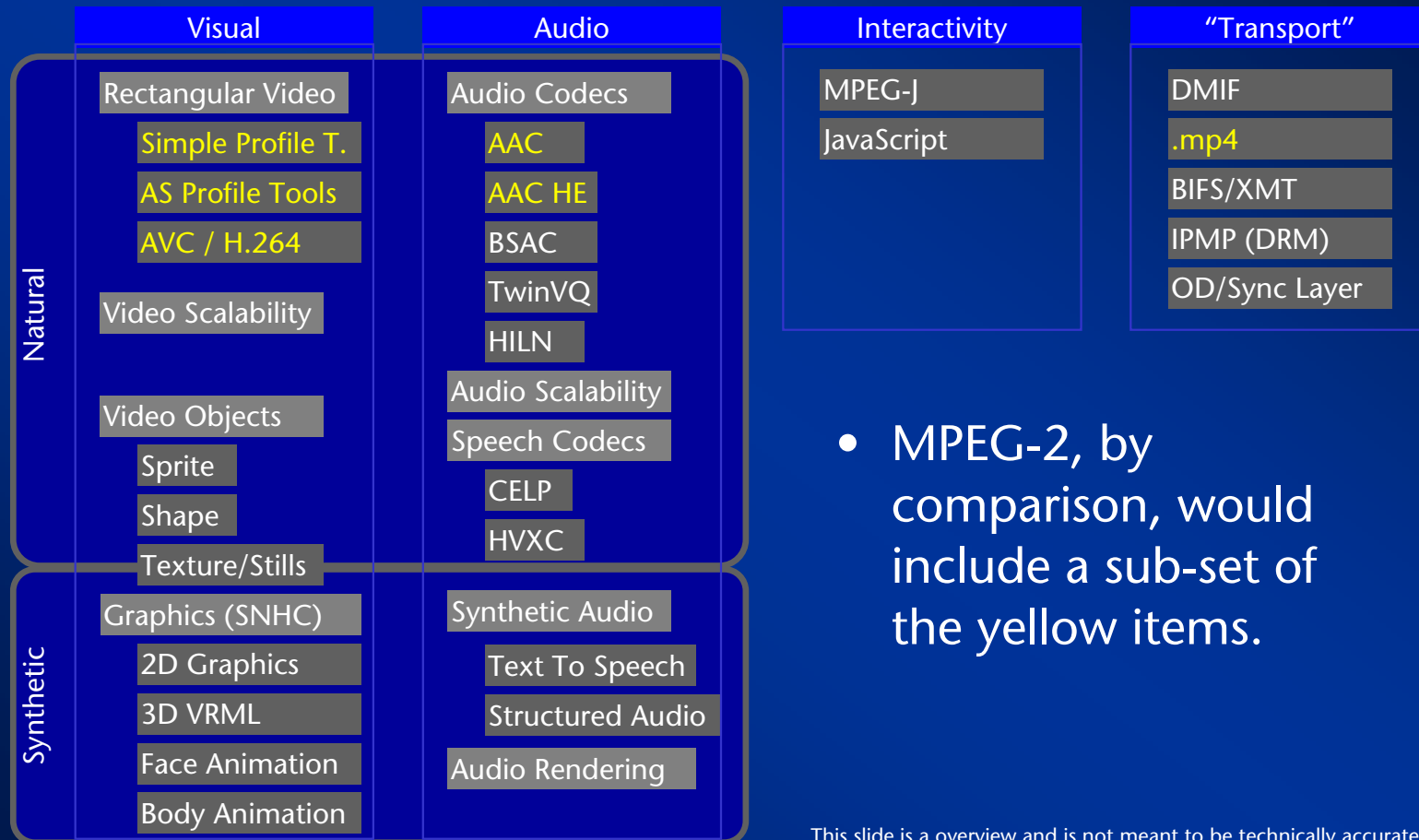
MPEG-4's Many Pieces

A Complex "All Things to All People" Toolkit

streamcrest℠

MPEG-4 – Designed before the Web to be All Things to All People

- Work began on MPEG-4 in 1993-1995

- More so than prior standards, MPEG-4 became an "all things to all people" project

- This has led to a "blind men describing the elephant" problem – different constituencies think of MPEG-4 as:

  - Improved audio and video compression with IP network delivery

  - High compression of video by encoding each object in a scene separately

  - Combining real ("natural") video objects in scenes with computer-generated ("synthetic") objects and audio.

  - Interactive and personalized content

  - Largest schism is between the "video" guys and the "CG/Virtual World" guys

streamcrest SM

One of the reasons MPEG-4 took so long is that it attempted to solve a bigger problem that just sending video and audio, and became an "all things to all people" project. The original idea was to use the ideas of MPEG-2 to send video over IP networks to PCs and mobile devices.

MPEG-4's Collection of Tools

Visual		Audio	Interactivity	"Transport"

Natural

Rectangular Video	Audio Codecs	
Simple Profile T.	AAC	
AS Profile Tools	AAC HE	
AVC / H.264	BSAC	
	TwinVQ	
Video Scalability	HILN	
	Audio Scalability	
Video Objects	Speech Codecs	
Sprite	CELP	
Shape	HVXC	
Texture/Stills		

Synthetic

Graphics (SNHC)	Synthetic Audio
2D Graphics	Text To Speech
3D VRML	Structured Audio
Face Animation	Audio Rendering
Body Animation	

Interactivity

- MPEG-J
- JavaScript

"Transport"

- DMIF
- .mp4
- BIFS/XMT
- IPMP (DRM)
- OD/Sync Layer

- MPEG-2, by comparison, would include a sub-set of the yellow items.

This slide is a overview and is not meant to be technically accurate

But, instead, what we got is this... So, what else is in MPEG-4?

Architecture of MPEG-4 extends previous standards to robust, scalable, layered objects

- Video Coding Techniques are adapted from MPEG-2 and H.263, with some improvements from new discoveries or judged now feasible with current hardware.

  – Each new improvement, or Tool, adds 1-10% to coding gain over MPEG-2 on certain material.

- In Contrast to MPEG-2's single picture, MPEG-4 also allows layers of irregularly-shaped objects

- MPEG-4 extends the (rarely used) scalability of MPEG-2 to allow each object to be scaled in spatial (screen) resolution or in temporal frame rate.

- MPEG-4's bitstream coding allows better recovery from transmission losses.

MPEG-2 – One Picture

MPEG-4 – Several Objects can be sent separately and combined in decoder

streamcrest℠

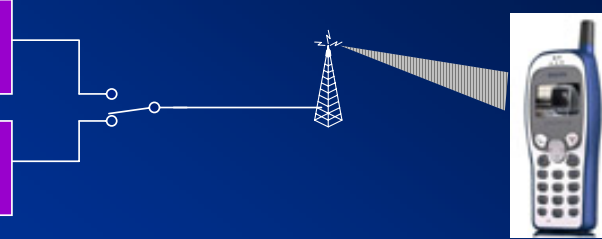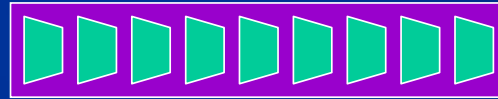For one thing, we can send layers of irregularly-shaped objects instead of one picture.
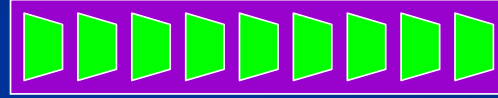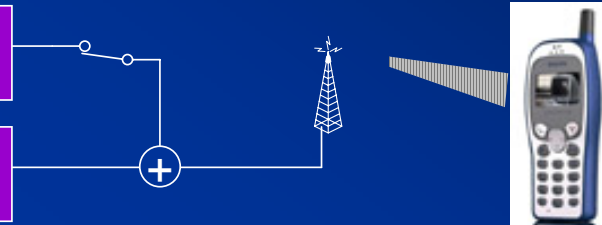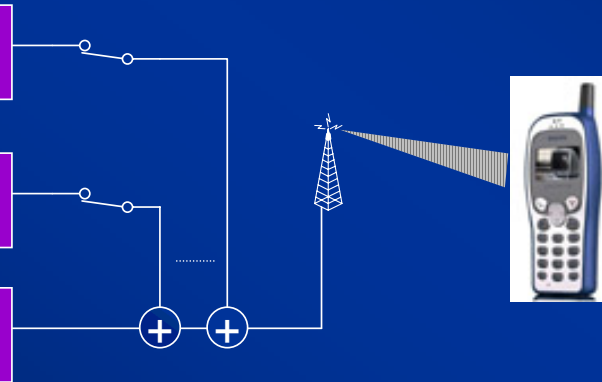
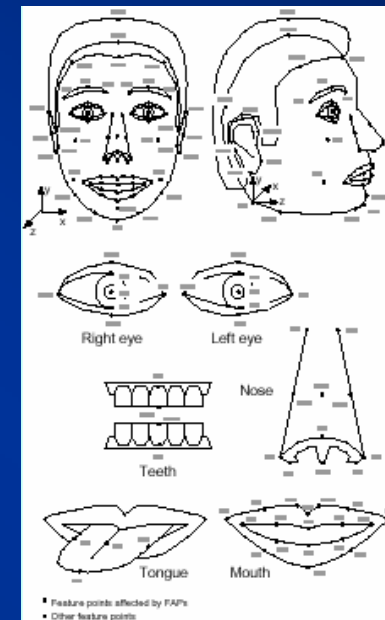For another, it offers several techniques for reducing the bitrate of a stream dynamically as the available bandwidth changes...

The Other MPEG-4

VRML meets the Video World

streamcrest℠

"The Other MPEG-4" Adds Synthetic Media Types to MPEG's Traditional Natural Media

- 2-D Mesh Objects

- 3-D Mesh Objects

- Face Animation

- Body Animation

- Structured Audio

Work on MPEG-4's synthetic media was started before Flash, Shockwave, SVG, and other proprietary formats were developed

It includes not just video, but 2 and 3-D graphics as well...

MPEG-4 is designed for interactivity, not just mashing couch potatoes

- Basic Interactivity is provided by a JavaScript-like programming language

- A modified form of Java, MPEG-J, is available for heavy-duty work

- The standard specifies (above the network level) a return channel for communicating back to a server

- Interaction is not just clicking hot spots. Users can move or change video and audio objects

streamcrest℠

Finally, it includes Java or JavaScript to allow the user to interact with the program.

The New MPEG-4:
The Advanced Video Codec

www.streamcrest.com

streamcrest℠

All this complexity slowed down MPEG-4 and allowed proprietary formats to be sometimes better on the fundamental job of encoding plain audio and video. So MPEG has responded with a new codec to fix that problem…

Different Names, Same Standard:

- JVT – (Joint Video Team)

- H.26L (also sometimes H.263++)

- MPEG-4 Part 10

- H. 264

- AVC (Advanced Video Coding)

- All refer to the same codec, standardized by joint ITU (H.261, H.263) and ISO (MPEG-1,2,4) group

- MPEG-4 Part 2 – Earlier codec in MPEG-4 Simple Profile, Advanced Simple Profile

- H.263 – Earlier ITU standard

streamcrest^{SM}

Because it was developed jointly by ITU and ISO, it has gone by a number of names, but the formal name today is AVC.

AVC offers high-performance back-to-basics video codec

- Perhaps because of its H.263 heritage, the AVC codec doesn't concern itself with objects, scalability, interactivity, and other MPEG-4 notions.

- It's just a very good codec for "ordinary" video just like we watch everyday.

- It is scalable from the smallest phone screens to digital cinema

- It takes 2-4 times more processing power to decode (perhaps 8 times for encoding)

- It offers the same quality at about 1/2 the MPEG-2 bitrate.

The AVC codec applies new tricks to encoding plain old regular video, without objects, and offers the same quality at about half the MPEG-2 bitrate.

Why is the AVC codec better?

- If there are no magic techniques, how do codecs improve performance?

- Coding efficiency is a tradeoff between:
  - **Cost** of hardware Gates or software MIPS needed to implement the encoder and decoder
  - **Latency** or delay in encoding and decoding the content
  - **Asymmetry** of encoding and decoding costs

- Progress in chip and computer power makes costly techniques feasible

- Additional experience through implementation and "Sun worship" allows algorithms to be tuned

AVC uses the same fundamental techniques I've presented, but it compresses better because it's now feasible to do more complex operations in hardware and software, and experience has allowed the experts to tune some tricks for better performance.

Video Codec Marketplace Competition and History

streamcrest℠

Will AVC succeed in the marketplace?

Video Codecs live in two different worlds today

- Hardware and Embedded World

- Needs Stable Standard

- Many competitive suppliers with interoperable products

- Dominant formats:
  - MPEG-2 for all forms of television and video (DVD, DBS, Digital Cable, professional video, etc.)
  - H.261/H.263 for Videoconferencing

- Software and Internet World

- Rapid Innovation and Performance Improvement

- Frequent Changes

- Proprietary closed standards

- Dominant formats:
  - Microsoft Windows Media
  - Real Networks
  - Macromedia Flash
  - Apple QuickTime

H.261/H.263 also has software implementations – such as AIM or NetMeeting

streamcrest℠

Today, open standards like MPEG-2 dominate the hardware world, where stable standards are important, but proprietary formats own software and Internet applications.

MPEG-4 AVC could replace all of these standards

- It is a competitive threat to proprietary standards because it offers similar levels of performance without their perceived proprietary control

- It's strength is that its stability and openness makes it attractive to semiconductor manufacturers to implement in hardware decoders

- It's weakness is that it is not available in a packaged solution – a ecosystem of support products has not developed yet.

3/5/2004

streamcrest℠

Strictly on its technical merits, AVC could replace all of these standards – it's as good or better than the proprietary ones without their proprietary control. But there is not a complete ecosystem of support products yet, and..

Patents plague some standards

- Compression saves money or enables new businesses, so inventors sometimes want a piece of the action.

- Unisys Patent on LZW Compression led to PNG format to replace GIF

- JPEG has remained royalty-free, but who's heard of royalty-bearing JPEG-2000?

- MPEG standards have traditionally involved royalties, while H.2XX series have not

- Interoperability and stability of standards have put them in a strong position in the past

Review Draft
Confidential Client Presentation © 2004 Streamcrest Associates

3/5/2004

streamcrest℠

We have a patent situation to contend with. In the past, some standards have been valuable enough for firms to collect patent royalties...

Royalty Costs of Competitive Formats

Format	MPEG-2	MPEG-4 SP, ASP		WMT		Real
	Unit	Unit	Cap	Unit	Cap	
Encoder	$2.50	$0.25	$1M	$0.20	$0.4M	$0.00
Decoder	$2.50	$0.25	$1M	$0.10	$0.8M	$0.25-$0.35
Codec	$2.50	$0.50	$2M	$0.25	$1M	n/a
Software	Same as HW	Same as HW		Same		Same
Server	Special terms for stat mux, transcoders, etc.	None		$18 - $900, $72 typ.		$500/CPU $13K/CPU for WMT, MPEG
Use	None	$0.02/hour or $0.25/user/year Cable/DBS STBs: $1.25 one-time		None		None
Packaged Media	$0.03 per movie	$0.02/hour		None		None

Some costs have been simplified or summarized here due to complicated license terms. www.streamcrest.com has more detailed summary slides of the MEPG-3 SP, ASP terms. http://www.microsoft.com/windows/windowsmedia/create/licensing.aspx explains the WMT terms. Real's $13K licensing cost is not published, but quoted by Real in a message on their BBS. Real decoder has certain floors and caps. Real rates from Helix program, may offer other terms

Including MPEG-2 - Two bucks goes into the patent pool every time you buy a DVD player. With MPEG-4, the patent holders decided to get greedy and charge not for the technology, but two cents an hour for the content. This is the main reason for the failure of MPEG-4 in many non-mobile markets.

MPEG-4, AVC being smothered by royalties

- AVC:

  - Two competing patent pools

  - Terms include charges for content as well as for encoders and decoders

- Proprietary Formats:

  - "Free" Implementations or modest charge

  - No content royalties

- Content royalties offend content industry
  A million or two is no problem, but 1×10^{-9} of revenues is

- Pools threaten future litigation against proprietary suppliers and customers

streamcrestˢᴹ

With the AVC codec, the matter is not decided, with two competing pools offering to license AVC patents and both continuing to charge content royalties.

Agenda

1	What's Mobile Video?	• The Early Years • Current Innovation	• Future Alternatives
2	Video Codecs and MPEG-4	• Why Compression? • Codec Evolution • MPEG-4: All Things to All People	• Scalability • The Other MPEG-4 • The AVC Codec • Royalties
3	Applications and Deployment Dilemmas	• Bandwidth Costs • Content Acquisition	• Operators vs. Services • Cultural Acceptance
4	Introducing the Experts	• Packet Video • Mindego	

3/5/2004

streamcrest℠

Mobile Video Applications and Deployment Dilemmas

streamcrest℠

The patent issues are not critical for mobile video, but there are business issues that are...

"Why should I stream a 3 MB video when I can get $3.00 for a 100-byte ring tone?"

(Strategist at Major U.S. Mobile Operator)

For one thing, the bandwidth to transmit video is still expensive, even with true 3G networks, and other applications may provide more short-term revenue with less expense.

Acquiring Content is Never Easy

- With voice and SMS services, content is self generated by the consumer

- As mobile applications become more information or entertainment-centric, operators must acquire (indirectly) interesting content.

- Content can be Created or Repurposed

- Creation Cost Examples (U. S.):
  - High-budget Feature Film - $500,000/minute
  - Soap Opera – $1,800/min
  - Network News Magazine - $1,000/min
  - Local (20th Market?) Newscast - $80/min (high-volume, variable)

Another issue is that attractive content is not easy to get. Practically speaking, the size of early audiences for mobile content means it will have to be re-purposed from other media instead of being created.

Operators Really Fear True Services

- Most operators come from a RBOC or PTT background – their focus has been on providing communications services

- Information and Entertainment services are foreign to them and scary

- Only innovators like DoCoMo have been willing to take the risks of developing new services – most operators want to wait until they are proven businesses before deciding to offer them

And most importantly, mobile operators aren't really comfortable with true entertainment or information services. Only leading edge innovators are currently comfortable with mobile video.

Touted Mobile Video Applications

- Music Video Download

- News/Sports Highlights

- Videophony

- Remote Monitoring (Traffic, Bar/Venue, Baby/Wife Webcam)

streamcrest℠

So what will these innovators offer?

The End

streamcrest^{SM}

Product and Company names listed here may be trademarks of their organizations

Product photographs and charts included here may be the property of third parties

This document is not for publication or further distribution without the permission of Streamcrest

Agenda

1 | What's Mobile Video?
- The Early Years
- Current Innovation
- Future Alternatives

2 | Video Codecs and MPEG-4
- Why Compression?
- Codec Evolution
- MPEG-4: All Things to All People
- Scalability
- The Other MPEG-4
- The AVC Codec
- Royalties

3 | Applications and Deployment Dilemmas
- Bandwidth Costs
- Content Acquisition
- Operators vs. Services
- Cultural Acceptance

4 | Introducing the Experts
- Packet Video
- Mindego

streamcrest ℠

Introducing The Speakers

- **Kathleen Peters** is Vice President, Device Solutions, for PacketVideo, working with silicon and device manufacturers worldwide. She has been a key executive at PacketVideo for more than four years.

- **Mikaël Bourges-Sévenier** is Mindego's founder and chief technical officer. He is a key contributor to the MPEG-4 Systems specification and one of the original developers of the MPEG-4 reference software.