Available online at www.sciencedirect.com
SCIENCE DIRECT
Future Generation Computer Systems 22 (2006) 929–939
FGCS FUTURE GENERATION COMPUTER SYSTEMS
www.elsevier.com/locate/fgcs

International real-time streaming of 4K digital cinema☆

Takashi Shimizu[a,*], Daisuke Shirai[a], Hirokazu Takahashi[a], Takahiro Murooka[a], Kazuaki Obana[a], Yoshihide Tonomura[a], Takeru Inoue[a], Takahiro Yamaguchi[a], Tetsuro Fujii[a], Naohisa Ohta[b], Sadayasu Ono[b], Tomonori Aoyama[c], Laurin Herr[d], Natalie van Osdol[d], Xi Wang[e], Maxine D. Brown[e], Thomas A. DeFanti[e], Rollin Feld[f], Jacob Balser[f], Steve Morris[f], Trevor Henthorn[g], Greg Dawe[g], Peter Otto[g], Larry Smarr[g]

[a] *NTT Network Innovation Laboratories, 1-1 Hikari-no-oka, Yokosuka, Kanagawa, 239-0847, Japan*
[b] *Research Institute for Digital Media and Content (DMC), Keio University, West Annex, 2-17-22 Mita, Minato-ku, Tokyo, 108-0073, Japan*
[c] *The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8656, Japan*
[d] *Pacific Interface Inc., 5351 Hilltop Crescent, Oakland, CA 94618, USA*
[e] *Electronic Visualization Laboratory, University of Illinois at Chicago, 851 S. Morgan St. Chicago, IL 60607-7053, USA*
[f] *Skywalker Sound, A Lucasfilm Company, PO BOX 3000, San Rafael, CA 94912-3000, USA*
[g] *California Institute of Telecommunications and Information Technology (Calit2), University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA*

Available online 30 May 2006

Abstract

This paper describes the world's first real-time, international transmission of 4K digital cinema and 4K Super High Definition (SHD) digital video at iGrid 2005, hosted at the California Institute of Telecommunications and Information Technology (Calit2) at the University of California, San Diego. Nearly six hours of live and pre-recorded 4K motion picture and audio content was streamed to iGrid in San Diego from the Research Institute for Digital Media and Content (DMC) at Keio University in Tokyo.

To implement this demonstration, several new technologies were introduced, including a prototype high-performance 4K compressed multicasting system called "JPEG 2000 Flexcast", and "Soundscape", a practical scheme for synchronizing audio and video transmitted from different locations over IP networks.

These iGrid 2005 demonstrations proved that it is now feasible to implement networked professional audio/video applications – production, post-production and distribution – even at 4K quality over IP networks up to 15,000 km long. The demonstrations also showed the new 4K motion picture technology being introduced for digital cinema can be usefully applied to other network applications such as remote telepresence, distance learning and scientific visualization.
© 2006 Published by Elsevier B.V.

Keywords: 4K SHD; JPEG 2000; Flexcast; Soundscape

1. Introduction

4K Super High Definition (SHD) images [1–3] have roughly 4000 horizontal and 2000 vertical pixels — offering approximately four times the resolution of the most widely-used HD television format, and 24 times that of a standard broadcast TV signal. 4K is a particularly significant new image format because it will be widely used for future digital cinema theatrical distribution under new specifications [4] proposed by Digital Cinema Initiatives, LLC, a consortium of the major Hollywood studios. At the same time, for scientists working with very large data sets in globe-spanning collaborations using distributed Grid computing, 4K digital media technology in combination with optical IP networks offers dramatic new visualization capabilities.

In the first public demonstration of 4K motion pictures in 2001, nearly all content had been played from local servers.

☆ This work is partially supported by Ministry of Internal Affairs and Communications, Japan.
* Corresponding author. Tel.: +81 46 859 3364; fax: +81 46 859 8585.
E-mail address: takashi.shimizu@acm.org (T. Shimizu).

0167-739X/$ - see front matter © 2006 Published by Elsevier B.V.
doi:10.1016/j.future.2006.04.001

The demonstrations at iGrid 2002 included 2K class real-time networked applications, such as Griz, TeraScope, TeraVision, HDTV over IP and so forth [5–8]. In 2002 and again in 2004, 4K-over-networks was demonstrated in Japan and in the USA using domestic networks [9]. However, the lack of a real-time compressor/decompressor (codec) capable of processing 4K motion pictures limited the use of 4K for "live" or "interactive" applications. Moreover, until recently, the difficulties of coordinating international networks with sufficient capacity and quality of service (QoS) for sustained "live" transmission of 4K limited the potential deployment of this new format in networked applications.

The main contribution of this paper is to prove that these historical barriers can now be overcome and a working end-to-end system can be implemented over long distances. Five organizations agreed to collaborate on a demonstration of real-time trans-Pacific 4K-over-IP at the iGrid 2005 event, drawing on the support of many others. The five primary organizers were: NTT Network Innovation Laboratories, Keio University Research Institute for Digital Media and Content (Keio/DMC), University of Illinois at Chicago Electronic Visualization Laboratory (UIC/EVL), University of California at San Diego California Institute for Telecommunications and Information Technology (UCSD/Calit2), and Pacific Interface Inc. For the Soundscape demonstrations, additional collaborators were invited to participate from the University of California at San Diego Department of Music, Skywalker Sound a Lucasfilm Company, San Francisco State University Institute for Next Generation Internet, and Youth Radio of San Francisco.

At iGrid 2005, for the first time, 4K was streamed "live" over IP Layer 2 and Layer 3 research networks between the USA and Japan spanning 15,000 km, at a sustained bit rate of 200–450 Mbps using real-time JPEG 2000 compression.

One of the biggest challenges was assembling and integrating one of the world's first end-to-end workflows for 4K digital motion pictures and multi-track digital audio over IP networks. Several leading manufacturers generously contributed essential 4K hardware components, several still in the prototype stage, such as 4K digital motion picture cameras, 4K digital projector, 4K codecs with multicast distribution system and 4K-capable real-time compute and storage combinations. Also, multiple 1 Gbps network paths had to be provisioned over JGN II, CAVEwave, Pacific Wave, GEMnet2 and CENIC.

iGrid 2005 demonstrations of 4K-over-IP succeeded in introducing to scientists at the workshop the potential of this new SHD visualization format for their applications. The demonstration also succeeded in showing the increased realism and interactive intimacy of 4K telepresence and the potential for improved distance learning using this technology. These demonstrations also proved the concept of using 4K-over-IP for commercial distribution of digital cinema at the highest quality, and then using the same infrastructure for delivery of so-called "Other Digital Stuff" in 4K to digital cinema theatres. Finally, these same systems were used to demonstrate the technical feasibility of working on digital cinema production and post-production at the highest 4K quality in collaboration with remote colleagues, even on the other side of the world, using the advanced IP networks emerging as part of the Global Lambda Integrated Facility (GLIF).



Fig. 1. Live greetings between UCSD and Keio Univ.

This paper describes the iGrid demonstration with a detailed explanation of the new technologies introduced: "JPEG 2000 Flexcast", a prototype high-performance 4K compressed multicasting system and "Soundscape", a practical scheme for synchronizing audio and video transmitted from different locations over IP networks. The paper is organized as follows. Section 2 describes the iGrid demonstration, the overall system and the network. The JPEG 2000 Flexcast system is described in Section 3, and the performance of the system is discussed in Section 4. Soundscape is described in Section 5 and Section 6 concludes this paper.

2. iGrid 2005 demonstration

2.1. Applications

The nearly six hours of content streamed from Keio/DMC to UCSD/Calit2 during iGrid 2005 included pre-rendered computer animations, materials shot with 4K digital motion picture cameras and digital still cameras, real-time computer-generated visualizations and digitally scanned 35 mm and 65 mm motion picture films. Applications demonstrated included: telepresence, distance learning, scientific visualization, cultural heritage simulation, digital cinema and other digital stuff (ODS) for entertainment.

Of these, several applications such as telepresence and distance learning were particularly challenging because their requirements for live interactive dialog between remote participants make them particularly sensitive to any potential problems of network quality of service (QoS), packet loss, latency, compression artifacts, and audio/video synchronization. For iGrid 2005, the 4K telepresence demonstration included live greetings between UCSD Chancellor Marye Anne Fox in San Diego, and Keio University President Yuichiro Anzai in Tokyo. It is shown in Fig. 1. Separately, 4K distance learning was demonstrated with an illustrated lecture on the Guttenberg Bible by Keio Professor Toshiyuki Takamiya, who used a combination of live lecturing on-camera, images of physical examples and close-up digital still photographs to show the

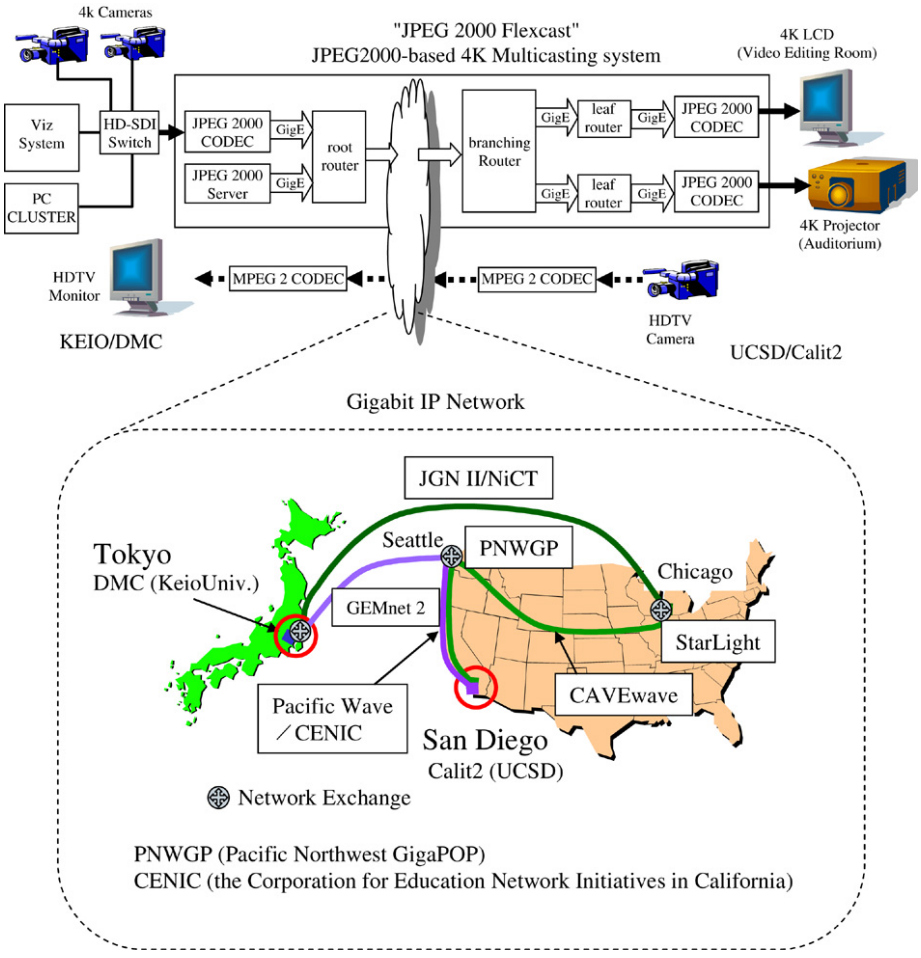


Fig. 2. Overall system.

fine points of Guttenberg's craft. Participants and audience alike commented on the high degree of "remote presence" achieved with 4K-over-IP in both demonstrations.

2.2. *System configuration*

These demonstrations required configuration of an end-to-end 4K system capable of live capture, pre-recorded playback, real-time JPEG 2000 codec, real-time multicast streaming, and large screen projection, connected together via a combination of video, audio and networking technologies.

Fig. 2 illustrates the overall system. At Keio/DMC, 4K digital motion picture cameras were connected via quad HD-SDI (SMPTE 292M) digital interfaces to the JPEG 2000 codec. (Each HD-SDI bit-stream had an uncompressed data rate of 1.485 Gbps.) The PC cluster that rendered Toppan's Nijo Castle and a visualization system were also connected to the JPEG 2000 encoder via quad HD-SDI signals which had been converted from quad DVI signals. The compressed JPEG 2000 bit-stream was transmitted to the gigabit IP network through the Flexcast root router. At UCSD/Calit2, the received stream was replicated by the branching router and distributed via IP networks to two locations: the auditorium and the video editing room. In each location, the Flexcast leaf router converted the stream coming from Keio/DMC to feed the JPEG 2000 codec appropriately, and each JPEG 2000 codec decompressed the streams to feed, respectively, a 4K digital projector in the auditorium and, simultaneously, a 4K liquid crystal display in the video editing room.

The system, as implemented, was capable of carrying six channels of uncompressed digital audio at up to 24-bit/48 kHz sampling resolution. Digital audio originating at Keio/DMC was input as discrete channels into the JPEG 2000 codec, which transmitted the uncompressed audio bit-streams in parallel with the compressed 4K visual bit-streams.

Various content was displayed for the iGrid audience over three successive days on a 30 ft wide by 15 ft high screen using a prototype 4K digital projector installed at the Calit2 auditorium. Given the architecture of the auditorium, audience members in the first 4 rows were seated within one picture height of the screen for an optimal 4K viewing experience. Even people sitting in the back row were less than three picture heights from the screen, still close enough to appreciate the extreme picture quality delivered by 4K.

In order to provide the capability for real-time interactive feedback between the sites in Tokyo and San Diego, additional network links were used to send a live HD video/audio stream from Calit2 to DMC to provide feedback to speakers and guests there. For this purpose, the 1.5 Gbps uncompressed output of a broadcast-quality HDTV camera in the Calit2 auditorium was

compressed by a stand-alone HDTV broadcast quality MPEG2 hardware codec, yielding a 50 Mbps RTP/UDP stream.

Flexibility and fine-tuning of a variety of audio assignment and diffusion scenarios was enabled by the use of a large format digital mixer. About 12 mixing presets were defined and utilized for the various 4K demos.

For example, in the 4K telepresence demonstrations, audio engineers in San Diego were able to direct and troubleshoot aspects of the Tokyo audio feed with accuracy and refinement, due to the clarity and size of the video image, and the immediacy of the low latency, bidirectional AV streams. Intelligibility and spatial accuracy were enhanced by assigning multiple microphone channels from the Tokyo feed across five discrete frontal channels of the 30 ft wide screen in the Calit2 auditorium in San Diego.

The Soundscape part of these demonstrations involved a live audio mix to match the picture in the auditorium, using 24 channels of uncompressed 24-bit/48 kHz digital music and sound effects originating from Skywalker Sound audio servers in the San Francisco Bay Area. The audio streaming from the Bay Area was synchronized over the IP networks with the 4K motion pictures streaming from JPEG 2000 servers at Keio/DMC.

2.3. Network

As shown in Fig. 2, the 4K real-time transmissions traveled from Tokyo to San Diego through a network of high-performance optical fiber links. The primary path crossed the Pacific on the JGN II/NiCT research network. This, in turn, connected through the optical switching hubs at StarLight in Chicago and the Pacific Northwest GigaPOP in Seattle to key American "LambdaGrids" CAVEwave, part of the US National LambdaRail, and Pacific Wave, inter-operating at the Ethernet (L2) layer. Additional L3 connectivity was achieved using NTT's trans-Pacific research network GEMnet2 that peered with Pacific Wave/CENIC at the Pacific Northwest GigaPOP. No difference in QoS was observed, except for the propagation delay.

3. JPEG 2000 Flexcast system

At 8 megapixels per frame, uncompressed streaming of 4K motion pictures requires a data rate greater than 6 Gbps. In many places though, the signal must be carried over 1 Gbps circuits. To do so efficiently, the iGrid demonstration utilized prototype JPEG 2000 codecs designed to compress and decompress 4K digital video in real time to 200–450 Mbps for direct connection to gigabit IP networks. These codecs, as well as related JPEG 2000 servers, were connected to prototype systems called "Flexcast" that enabled multicast delivery of 4K video and audio over traditional unicast networks.

3.1. JPEG 2000 real-time codec

The prototype JPEG 2000 codec [10] can perform real-time compression and decompression of up to 250 megapixels per second, yielding a compressed data rate of up to 500 Mbps. The codec consists of a PC running Linux with a Gigabit Ethernet network interface, and PCI-X JPEG 2000 codec boards with four JPEG 2000 processing chips. This codec can process images up to 4096 × 2160 pixel resolution with up to RGB (4:4:4) 36-bit color at up to 24 frames per second in real time. In addition to the so-called 4K digital cinema distribution master (DCDM) image format described above, the codec also works with smaller frame sizes of 3840 × 2048 pixels, 3840 × 2160 pixels (Quad HDTV mode), 2048 × 1080 (2K DCDM mode) and 1920 × 1080 (HDTV mode). It can also support a frame rate of 30 fps (TV mode) at all these resolutions when using YPbPr (4:2:2) color encoding.

The codec manages the accompanying audio data via a PCI digital sound card that can output 6-channel, 24-bit/48 kHz digital audio via ADAT digital I/O. The audio data is not compressed to preserve the highest sound quality to accompany the 4K large-screen digital imagery applications. The total audio bit rate is only 7 Mbps, which is negligible compared to the data rate of the video signal.

This codec can be connected to input/output devices using four SMPTE 292M HD-SDI Dual Links via TMDS/HD-SDI signal converters. However, a Quad-Dual Link digital interface for 4096 × 2160 RGB 4:4:4 picture format has not yet been defined. Therefore, the output from this codec is currently limited to 3840 × 2160 pixels YPbPr 4:2:2, so-called Quad HDTV mode.

3.2. JPEG 2000 server

The JPEG 2000 server can store and play-out motion picture sequences that have been pre-compressed using JPEG 2000, starting with 16-bit TIFF files up to 4096 × 2160 resolution. The server consists of a PC running Linux with dual CPUs (Xeon, 3.0 GHz), a high-performance Software-RAID (Serial ATA disk 250 GB × 7, RAID 0, XFS file format), and a Gigabit Ethernet (GigE) network interface. Because JPEG 2000 is an intra-frame codec, the server was designed to handle clips of any length as a sequence of still images. A server program was developed that receives a data transfer command, reads the data from the RAID and writes the data periodically to the GigE network interface. The server has a particularly large shared buffer in order to smooth-out any variations in disk read speed so as to maximize data transmission speed.

3.3. Flexcast

"Flexcast" [11–13] is a mechanism to provide large scale multicast distribution services over IP networks for all types of streaming media, scalable up to SHD 4K motion pictures with synchronized audio. In this demonstration, it is used for delivering 4K content to two distinct output devices: 4K projector and 4K LCD. Because Flexcast uses only unicast UDP packets, it can work in networks that do not support IP multicast routing. Using Flexcast, a multicasting infrastructure can be created incrementally by adding Flexcast functions to existing unicast IP infrastructures. This feature is extremely promising for 4K applications. Because in the early stage,

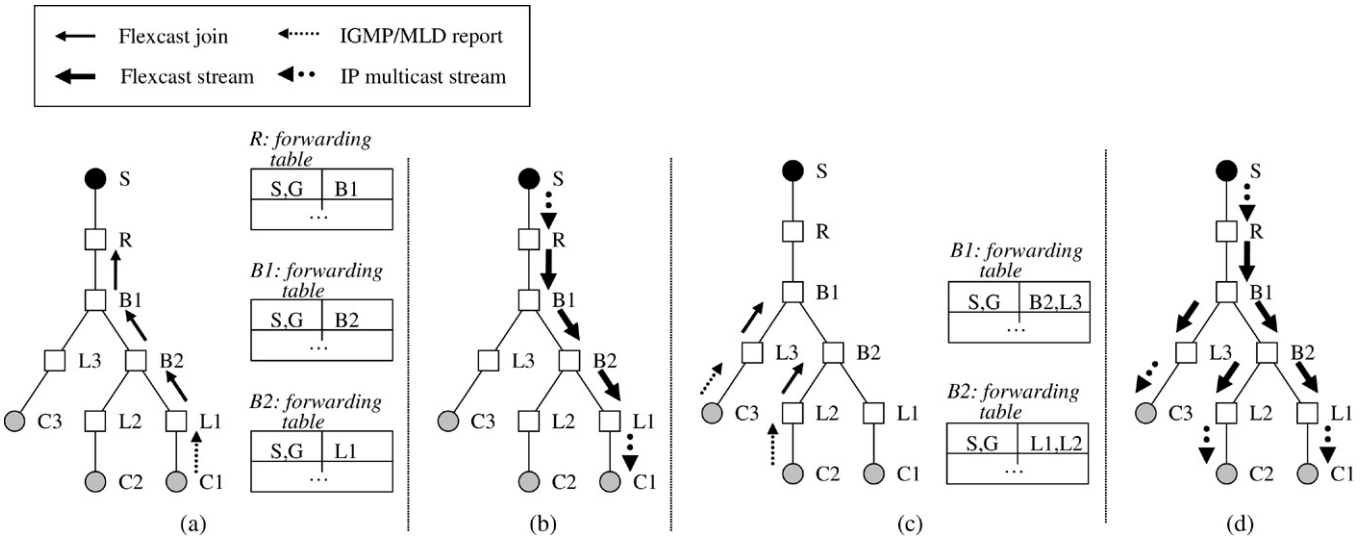


Fig. 3. Flexcast operation.

potential users of such applications tend to be sparse and isolated, where large-scale initial deployment may not be a reasonable choice for service providers.

In Flexcast, join packets construct a tree-structured delivery path. Fig. 3 shows an example. A router located in a segment to which clients are connected, is called a "leaf" router. A router located at the ingress point of the segment to which a multicast source is connected, is called a "root" router. A router located on the path between leaf routers and root router, is called a "branching" router.

Suppose that clients C1, C2, and C3 are trying to receive packets from multicast source S. At first, C1 emits an IGMP membership report or a MLD listener report, which contains the address of S and the group address G. Here, Flexcast is designed to be compatible with IP multicast, as it uses SSM channel abstraction to identify a tree and controls recipients by IGMP/MLD. When leaf router L1 receives the report, it sends a join packet to S. Each join packet contains channel information (S, G). Branching router B2, which is located on the path from L1 to S, intercepts the join packets and registers L1 in its forwarding table. B2 then sends a join packet to S. In the same manner, the branching router B1 registers B2 in its forwarding table, and sends a join packet to S.

When root router R intercepts the join packet, it registers B1 in its forwarding table similarly. At this time, the forwarding tables of the routers are shown in Fig. 3(a). When R receives a multicast packet addressed to G from S, R encapsulates the packet into a stream packet and sends it to B1. The stream packet is forwarded by B1 and B2, and finally arrives at L1. The leaf router L1 receives the stream packet and decapsulates the original multicast packet, which is released to the segments in which C1 exists, as shown in Fig. 3(b).

When C2 and C3 send out reports, L2 and L3 receive them, respectively, and send join packets to S similarly. As a result, L2 and L3 are registered in the forwarding tables of B2 and B1, respectively, as illustrated in Fig. 3(c). The multicast packet from S is encapsulated by R into a stream packet, which is delivered to each leaf router. Finally, L2 and L3 decapsulate the original multicast packet and send it out to their clients, as described in Fig. 3(d).

Suppose that B2 does not support Flexcast. Join packets sent by L1 and L2 are intercepted by B1, which registers L1 and L2 in its forwarding table. The stream packet is replicated at just B1 and directly delivered to L1, L2, and L3. In this way, even if some routers do not support Flexcast, a multicast stream can be delivered to each client. By replacing B2 with a Flexcast branching router, redundant traffic on a link between B1 and B2 is removed. The more Flexcast branching routers there are, the less redundant traffic there is. This capability for incremental deployment allows initial implementation without significant investment or disruption of existing network infrastructure, without limiting subsequent expansion options. Note that Flexcast is designed to work over both IPv4 and IPv6 networks.

4. Performance of JPEG 2000 Flexcast

4.1. Multicasting

Both the JPEG 2000 real-time codec and the JPEG 2000 server transmitted 5 UDP flows for efficient 4K streaming. The sending rate of the flows was controlled by inserting an adequate busy-wait between each UDP packet sent, in order to reduce the peak rate of the flows.

Fig. 4 shows the time series plot of multicast traffic measured at the input port of two leaf routers in Calit2. The horizontal axis denotes the time in 200 ms units, and the vertical axis denotes the total amount of multicast traffic in gigabits per second. Each point was the average of 200 consecutive measurements that were taken every 1 ms. The mean rate averaged over 1 s period was about 500 Mbps. The two streams are indicated as a solid line in the upper half and a dotted line in the bottom half, respectively. As can be seen, the shapes are almost the same. This shows that the performance of the PC-based Flexcast branching router used for iGrid 2005 is sufficient to replicate the JPEG 2000 stream for two decoders.

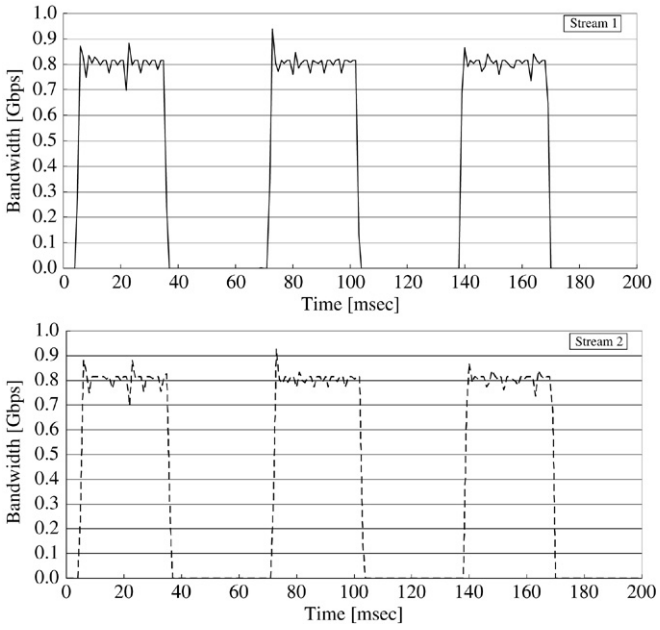


Fig. 4. Measurement.

4.2. Forward error correction

Both the codec and the server implemented forward error correction (FEC). At the server side, compressed data of a frame is divided into equal-sized packets of a specified size. A data chunk consists of specified numbers of packets. The last data of a frame is padded to fill the size of a data chunk. At each data chunk, horizontal parity was generated to create a parity packet. The server adds a header to each packet and sends these packets in interleaved fashion. The JPEG 2000 decoder receives these packets and stores them in an adequately sized receive buffer. Before writing the data into the JPEG 2000 hardware decode buffer, the decoder checks the arrival of each packet and tries to recover any lost data packets using the parity packet of that data chunk. If one packet in a data chunk is lost, the decoder can recover the entire data chunk using the data on the parity packets. Otherwise, it takes advantage of the natural coherence in sequences of motion pictures and copies a comparable data chunk from the previous frame (still held in the decoder's receive buffer). The interleaved packet ordering used in this system is designed to reduce the probability of multiple losses.

Throughout the iGrid demonstrations of 4K-over-IP, packet loss occurred frequently. It may be caused from the mixture of packet clumping and temporal concentration of traffic at some switches in the path. However, the FEC function significantly helped prevent data loss. Adding 14% parity data for FEC to one 4K compressed data stream of 200–400 Mbps, we observed 243 recovered frames and 18 lost frames out of 35,777 frames total. Using FEC, frame error rate was improved from 0.73% to 0.05%. Most of the losses were recovered using the 14% FEC rate, though there were the cases that could not be recovered. This is thought to be primarily caused when the burstiness of the packet loss was more severe than the FEC tolerance level achieved by interleaving.

4.3. Round trip latency

Total system configuration between Tokyo and San Diego included many potential sources of latency. In order to assess overall system performance for interactive 4K telepresence in terms of the critical factor of round trip latency, a simple method to estimate time lag was used, as follows. A large LED clock showing "hours:minutes:seconds:hundreds of a second" was placed at Keio/DMC so that it would appear in all scenes shot using the 4K cameras. As previously described, the 4K camera output from Keio/DMC was streamed to Calit2, and projected on the screen in the auditorium. The screen image was shot by an HDTV camera at Calit2, streamed back to Keio University, and displayed on an HDTV monitor. The clock in the HDTV monitor and the real clock were compared to get a rough measurement of overall round-trip system latency. The difference in the "real time" and the "delayed time" averaged slightly more than 1 s. Since the network-level round trip time was about 130 ms, further reduction of the overall latency is possible by optimizing each component of the system.

5. Soundscape

Soundscape is the name given to the technology and performance art intended to demonstrate next-generation multi-channel networked audio streaming for interactive applications like telepresence or distance learning, for live audience screenings as might occur in a digital cinema theatre, for scientific visualization, and for new audio post-production techniques that could be applied to digital cinema. In particular, the ability to synchronize geographically separated AV streams enables practical and powerful "distance" production and post-production solutions that have been unattainable in the past, due to computation and configuration inefficiencies, difficulties of assuring network QoS, excessive network latency, and bandwidth limitations.

Such capabilities are increasingly important because the very highly-paid creative people who make modern cinema want access to specialized facilities like the enormous sound recording studios at Skywalker Sound and advanced 4K screening rooms via networks so that they do not have to travel, but only if the networks can enable full quality, full performance and full creative control, just as if they were physically in the same location. The idea that this can be made feasible using fast optical networks is the founding principle of the CineGrid initiative, of which iGrid 2005 was the first demonstration. To make the Soundscape demo have real meaning to the professionals at Skywalker Sound, we targeted the highest quality audio and visual formats, such as might be used to post-produce digital cinema in Hollywood, streamed in real-time over IP networks, from distant audio and video sources, synchronized on-the-fly such as would be normally done in a "live" final multi-track audio mixing session with synchronized projection of the motion picture.

A significant initial challenge for Soundscape was to find an audio-over-IP "packetizing" scheme robust enough to accommodate the required bandwidth and interactive

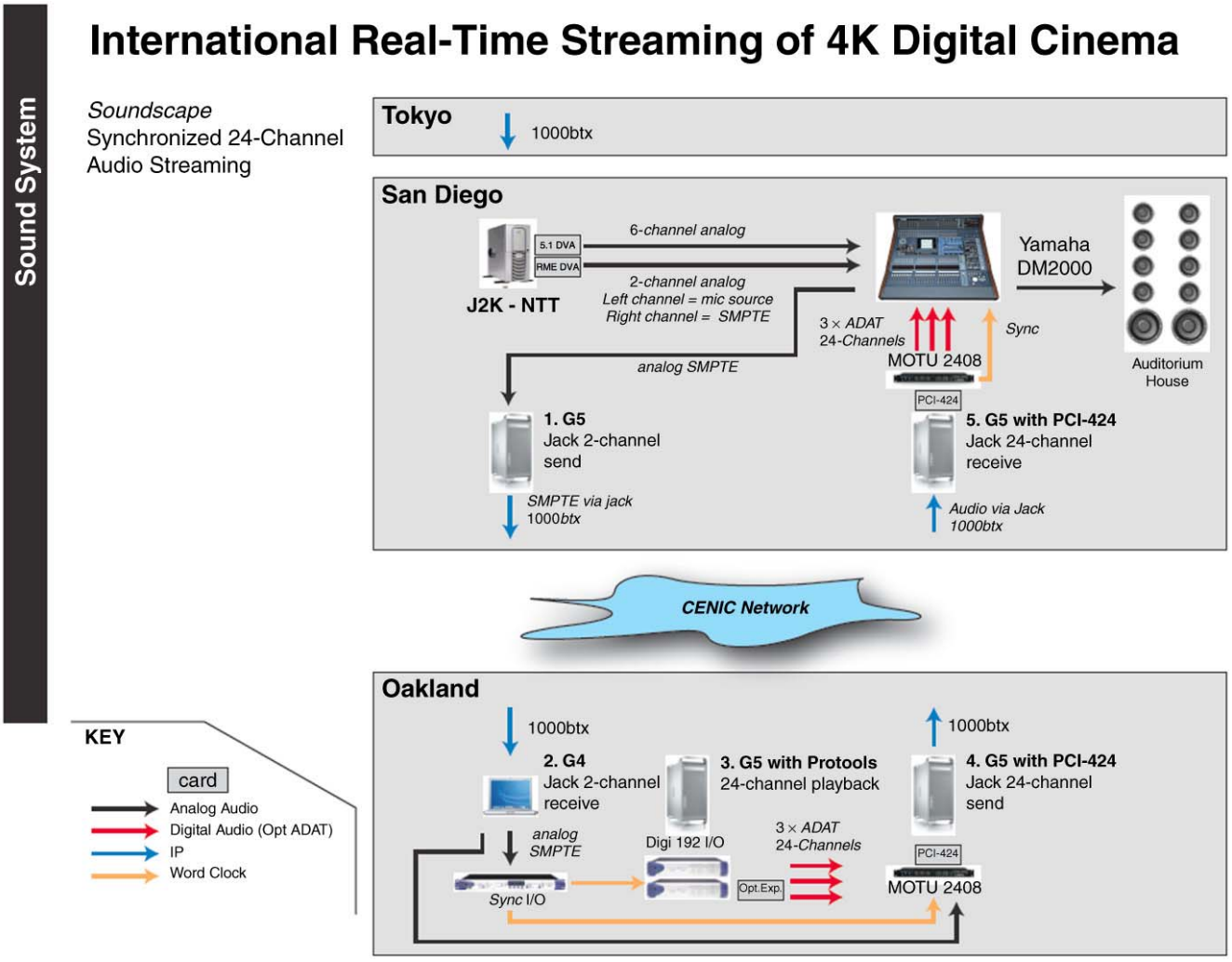


Fig. 5. Soundscape system diagram.

production flexibility. As a practical matter, this meant solving latency, synchronization, channel-count, and remote-control issues, plus finding a way to switch audio system configurations quickly to allow for the variety of diffusion and patching formats used in the different 4K demonstrations.

Soundscape was designed to meet the following criteria:

- Preserve maximum audio quality, avoiding audio compression if possible, running at a minimum of 48K sample rate with 24-bit precision and delivering up to 24 channels of sample synchronous audio
- Maintain robustness of packet integrity with low latency throughout and correct for any detected latency with sub-frame accuracy.
- Utilize low-cost, generic networking, audio streaming and production technologies, using software rather than hardware-based solutions whenever possible.
- Allow for scalability of channel count and diffusion assignments in anticipation of changing surround sound delivery conventions such as 10.2 and above (the Calit2 auditorium can accommodate up to 22.2 surround).
- Incorporate standard AV synchronization strategies (specifically SMPTE Time Code via IP) allowing audio to "chase" video for flexibility and interactivity in post-production, even when the audio and video are streamed over different IP networks from different sources.

A solution for audio-over-IP packetizing was found in the open source, inter-application software package known as Jack, a multiplatform audio pipe that we chose to run on Apple Mac G5 personal computers. Jack is similar in concept to commercial packages such as Propellerhead's Rewire, but with an added feature that allows IP network sockets to pass multi-channel audio in sample-synchronous blocks [14]. The industry-standard ProTools audio server was configured for multi-channel output in time-code "chase" mode. ProTools, however, is not "Jack-friendly" and thus a bridge between ProTools output and IP input was established using a second Mac G5 on the server site, configured with a MOTU 2408/PC424 interface with three ADAT optical 8-channel bundles. Jack software allowed internal bus patching of device inputs (in this case MOTU 2408 inputs) to networked target IP addresses. Although Jack appears to be limited to

16 audio channels of network node connectivity, a simple solution was found in assigning a second IP address to the receiving (networked audio target) G5. Jack recognized this second receiving address and eight additional channels were assigned to yield the desired total of 24 IP-packetized audio channels. The 23 main audio channels were then output via a second MOTU 2408 and patched via ADAT into the Calit2 auditorium's sound system. This solution was set up and tested in various networking topologies and simulations at Skywalker Sound and in UCSD's computer music studios and in Calit2's new Spatial Audio Lab.

The final production system diagram is shown in Fig. 5. A solution for the problem of parallel audio and video streaming from two distant sources was devised by "striping" a SMPTE time code track on several of the 4K animations to be streamed from Tokyo. Upon arrival in San Diego, this SMPTE time code signal would be decoded and analyzed on a Mac G4 laptop, establishing a local reference TC clock, then regenerated and re-routed to Oakland, where it would drive a ProTools multichannel session. The time code stream would be copied into a channel of the ProTools audio stream and then the composite of 23 main audio channels plus one SMPTE time code channel was encapsulated in IP via Jack, and sent back to San Diego via CENIC's 1 Gbps HPR link. In San Diego a Mac G5 translated IP back into audio (using Jack in receive mode) and then routed the time code channel coming from Oakland back to the Mac G4 laptop for comparison with the original time code coming from Tokyo. The "round trip" time differential was calculated and the San Diego-to-Oakland time code was accordingly adjusted forward in relative time, then reflected back to Oakland, where it "slipped" the audio server ahead to compensate for the latency differential. A Max/MSP applet was written to handle these latency adjustments. All time code analysis and adjustment occurred quickly during a few seconds of silent "pre-roll". Round-trip latency in the San Diego–Oakland loop was, estimated to be in the 21 ms range.

6. Conclusion

Taken altogether, the iGrid 2005 demonstrations dramatically showed that it is now feasible to implement a wide variety of audio-video applications at the highest quality over IP networks with multicasting, using new approaches to packet-loss counter-measures such as FEC, latency reduction and compensation techniques, high-speed packetizing of both audio and video, and synchronization of audio and video streams arriving at a performance space from different locations via different IP networks. And that the 4K motion picture format can be productively applied to networked telepresence, distance learning, scientific visualization and entertainment of various types.

The JPEG 2000 Flexcast system described in this paper was shown to have the capability to multicast 4K SHD moving images using 1 Gbps allocated bandwidth via both Layer 2 and Layer 3 network connections that would not otherwise support multicast. In order to reduce the bandwidth without significantly degrading picture quality, real-time JPEG 2000 compression was successfully used at various bit rates up to 450 Mbps.

The Soundscape demonstrations proved that AV synchronization could be maintained with no noticeable degradation once latency compensation was achieved. And further, that it is now technically feasible to build a professional quality multichannel streaming audio system that can be slaved to the transport control of a remote video system. The Jack solution for IP packetizing was found to be robust when clocking and set-up conventions were understood and carefully observed.

Acknowledgements

Generous support in the form of equipment loans, technical cooperation and contributions of content was gratefully received from many people and organizations in Japan and USA to enable these experiments. NTT Network Innovation Labs provided the 4K JPEG 2000 codecs, the JPEG 2000 servers and the 4K Flexcast systems. Sony Electronics provided the 4K SXRD projection system. Olympus Corporation loaned two of its SH-880TM prototype 4K digital motion picture cameras to Keio/DMC to give iGrid 2005 the highest-resolution live telepresence system in the world. SGI loaned its new Silicon Graphics Prism™ visualization system capable of playing back uncompressed 4K digital cinema clips in real time from the SGI® InfiniteStorage RM660 system.

A complete list of contributors includes: Digital Cinema Technology Forum (Japan), Digital Cinema Consortium of Japan, National Institute of Information and Communications Technology (Japan), Olympus Corporation, Toppan, Sony Electronics, Inc., Silicon Graphics, Inc., SGI Japan, ASTRODESIGN, Mitsubishi Electric, Yamaha Corporation of America, University of Illinois Urbana Champaign National Center for Supercomputer Applications, University of Southern California School of Cinema-TV, San Diego State University, Tokyo University of Technology Creative Lab, Tatsunoko Production Co., Ltd., The Pixel Farm, DALSA, Miranda, BAPS Swaminarayan Sanstha, Skywalker Sound a Lucasfilm Ltd. Company, San Francisco State University Institute for Next Generation Internet, Youth Radio, JGN2/NiCT, CAVEWave, PacificWave, CENIC, StarLight, Pacific Northwest GigaPOP, and University of California Office of the President.

References

[1] S. Ono, N. Ohta, T. Aoyama, All-digital super high definition images, Signal Processing: Image Communication 4 (4–5) (1992) 429–444.
[2] N. Ohta, S. Ono, Super high definition image communication — applications and technologies, in: Proc. Fourth International Workshop HDTV and Beyond, 1991, pp. 301–308.
[3] S. Ono, J. Suzuki, Perspective for super-high-definition image systems, IEEE Communications Magazine (1996) 114–118.
[4] Digital Cinema Systems Specifications, Digital Cinema Initiatives, LLC Technology Committee, 2005.
[5] L. Renambot, T. van der Shaaf, H.E. Bal, D. Germans, H.J. Spoelder, Griz: Experience with remote visualization over an optical grid, Future Generation of Computer Systems 19 (6) (2003) 871–881.
[6] C. Zhang, J. Leigh, T.A. DeFanti, M. Mazzucco, R. Grossman, Terascope: Distributed visual data mining of terascale data sets over photonic networks, Future Generation of Computer Systems 19 (6) (2003) 935–943.

[7] R. Singh, J. Leigh, T.A. DeFanti, F. Karayannis, Teravision: A high resolution graphics streaming device for amplified collaboration environments, Future Generation of Computer Systems 19 (6) (2003) 957–971.

[8] N.T. Karonis, M.E. Papka, J. Binns, J. Bresnahan, J.A. Insley, D. Jones, J.M. Link, High-resolution remote rendering of large datasets in a collaborative environment, Future Generation of Computer Systems 19 (6) (2003) 909–917.

[9] T. Yamaguchi, D. Shirai, T. Fujii, T. Fujii, SHD digital cinema distribution over a global high-speed network: Internet2, NTT Technical Review 1 (5) (2003) 10–15.

[10] T. Yamaguchi, M. Nomura, K. Shirakawa, T. Fujii, SHD movie distribution system using image container with 4096 × 2160 pixel resolution and 36 bit color, in: Proc. ISCAS 2005 IEEE International Symposium on Circuit and Systems, 2005, pp. 5918–5921.

[11] S. Tani, T. Miyazaki, N. Takahashi, Adaptive stream multicast based on IP unicast and dynamic commercial attachment mechanism: An active network implementation, in: Proc. IEEE IWAN2001, 2001.

[12] T. Inoue, S. Tani, K. Ishimaru, S. Minato, T. Miyazaki, Wide-area multicasting based on flexcast: Toward the ubiquitous network, in: Proc. APSITT2003, 2003.

[13] T. Inoue, S. Tani, H. Takahashi, S. Minato, T. Miyazaki, K. Toyoshima, Design and implementation of advanced multicast router based on cluster computing, in: Proc. ICPADS2005, 2005.

[14] See, http://www.jackosx.com/.



Takashi Shimizu is a senior research engineer of NTT Network Innovation Laboratories. Shimizu received a Dr.E degree in electrical engineering from Tokyo Institute of Technology in 1996. Since then, he has been engaged in researching IP traffic management mechanisms for distributed interactive applications. In 2004, he spent a year as a visiting scholar at Stanford University. His interests include terabit-level hardware design and photonic network management.



Daisuke Shirai received his B.Eng. degree in electrical engineering and M.Eng. degree in computer science from Keio University, Tokyo, Japan, in 1999 and 2001, respectively. He joined NTT Network Innovation Laboratories, Yokosuka, Japan in 2001 and has been researching super high definition digital cinema systems and their transmission technology. His interests include terabit-level image transmission and scalable video distribution. He is a member of IEICE of Japan.



Hirokazu Takahashi is a researcher of NTT Network Innovation Laboratories. He received the B.E. and M.E. degrees in electrical engineering from Nagaoka University of Technology in 2000 and 2002, respectively. After joining in 2002, he has been researching multicasting technologies for real-time applications. His current interests include reliable multicasting techniques.



Takahiro Murooka is a senior research engineer of NTT Network Innovation Laboratories at Kanagawa, Japan. He received the B.S degree in physics from the Saga University, Saga, Japan in 1989, and the Ph.D. degree from Kyoto University in 2006. Since joining NTT in 1989, he has been engaged in research on system design methodology and FPGA CAD system. His research interests are in architecture of programmable telecom-node and high-level system design method. Dr. Murooka is a member of IEEE, ACM, IPSJ, and IEICE.



Kazuaki Obana is a research engineer of NTT Network Innovation Laboratories. He received his M.E. degree in electrical engineering from Waseda University in 1997. His current technical interest includes applications on broadband-ubiquitous networks.



Yoshihide Tonomura received the B.S. and M.S. degrees in electronics engineering from Nagaoka University of Technology in 2002 and 2004, respectively. He joined NTT Network Innovation Laboratories in 2004. His research interest focuses on image processing theories and applications.



Takeru Inoue is a researcher of NTT Network Innovation Laboratories. He has been engaged in R&D on mobile networking, multicast communication, and multimedia conferencing system. From 2005, Inoue is also a Ph.D. student in Communications and Computer Engineering, Graduate School of Informatics, Kyoto University. He received the research awards of the Institute of Electronics, Information, and Communication Engineers of Japan (IEICE) Information Network Group in 2002 and 2005. He also received the best paper award from Asia–Pacific Conference on Communications in 2005.



Takahiro Yamaguchi received his B. Eng., M. Eng. and Ph.D. degrees in Electronic Engineering from the University of Electro-Communications, Tokyo, Japan, in 1991, 1993, and 1998, respectively. He joined NTT Optical Network Systems Laboratories, Yokosuka, Japan in 1998 and has been researching super high definition imaging systems and their digital cinema applications. He is currently a senior research engineer in NTT Network Innovation Laboratories.



Tetsurou Fujii received the B.S., M.S., and Ph.D. degrees, all in electrical engineering, from the University of Tokyo, Tokyo, Japan, in 1979, 1981, and 1984 respectively. In 1984, he joined Nippon Telegraph and Telephone Public Corporation (now NTT), Japan. He stayed at the University of California, Berkeley, from 1988 to 1989, as a visiting researcher. He has been engaged in research on adaptive signal processing, image signal processing and transport processing. His current interest is Super High Definition image processing and its application to digital cinema. He became an Executive Manager of Media Networking Laboratory, NTT Network Innovation Laboratories in 2005.

Major awards Achievement Award of IEICE (The Institute of Electronics, Information and Communication Engineers), in 2004. Best Technical Achievement Award of AMD (Association of Media in Digital), 2005.



Naohisa Ohta, Ph.D. joined NTT Laboratories in 1978 where he researched and developed signal processing algorithms for audio-video communications and highly parallel DSP systems and architectures. He joined Sony in 1998 and directed R&D on AV transmission with QoS for real-time applications, scalable coding for high-quality digital cinema, extra reality video creation technologies, and personalization technologies, etc. as a President of Broadband Applications Laboratories from 2002 to 2004. He is currently a Professor of Research Institute for Digital Media and Content of Keio University, Japan and leads a project including 4K digital cinema technologies, networked content sharing and digital archiving. He is an IEEE Fellow and currently serving as a Member at Large of Board of Governors in IEEE ComSoc.



Sadayasu Ono received his B.E., M.E. and Ph.D. degrees from Keio University, Tokyo, in 1971, 1973 and 1976, respectively. He is currently a professor of Research Institute for Digital Media and Content, Keio University, and was formerly a research fellow at NTT Laboratories. His research interests include digital signal processing, digital image processing, and digital cinema. Hollywood's standard for digital cinema, DCI (Digital Cinema Initiatives LLC) 's specification V1.0, 2005, is based on his research results, and are known as Super High Definition Images.



Tomonori Aoyama received the B.E., M.E. and Dr. Eng. from the University of Tokyo, Japan, in 1967, 1969 and 1991, respectively. Since he joined NTT Public Corporation in 1969, he has been engaged in research and development on communication networks and systems in the NTT Electrical Communication Laboratories. From 1973 to 1974, he stayed in MIT as a visiting scientist to study digital signal processing technologies.

In 1994, he was appointed to Director of the NTT Opto-Electronics Laboratories, and in 1995 he became Director of the NTT Optical Network Systems Laboratories. In 1997, he left NTT, and joined the University of Tokyo. In April 2006, he moved to Keio University, and is currently Professor of Research Institute for Digital Media and Content, Keio University.

Dr. Aoyama is an IEEE Fellow and an IEICE Fellow. He is currently serving as Vice-President of IEICE. He serves as Chair of the Photonic Internet Forum in Japan, the Digital Cinema Consortium of Japan, the Digital Cinema Technology Forum and Vice-chair of the Ubiquitous Networking Forum.



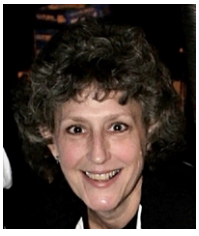
Laurin Herr is founder and president of Pacific Interface Inc. (PII), an international consulting company based in Oakland, California. He has also held senior management positions at several leading Silicon Valley manufacturers of computer/video peripherals. He has served as an advisory member of the Digital Cinema Consortium of Japan since its inception in 2001, and is a member of SMPTE and ACM SIGGRAPH. Herr received his Bachelor of Arts degree from Cornell University and pursued additional graduate studies at Cornell and at Sophia University in Tokyo.



Natalie Van Osdol is Vice President of Pacific Interface, Inc (PII), an international consulting company based in Oakland, California. Van Osdol is also a founding partner of Compression Technologies, Inc., a company dedicated to the development and licensing of digital video compression tools. She attended Sophia University in Tokyo and UCLA.



Xi Wang is currently a visiting post-doctoral research associate with the Electronic Visualization Laboratory, Computer Science Department at the University of Illinois at Chicago, USA. He received the B.S. degree from Doshisha University, Japan in 1998, and the M.E. and Ph.D. degrees from the University of Tokyo, Japan, in 2000 and 2003. His research interests include the design, development and implementation of innovative networking technologies (hardware, software, grid middleware), and the overall design of a seamless global optical infrastructure.



Maxine D. Brown is an Associate Director of the Electronic Visualization Laboratory (EVL) at the University of Illinois at Chicago (UIC). Currently, she is co-principal investigator of the NSF international networking TransLight/StarLight award, project manager of the NSF-funded OptIPuter initiative, a member of the PRAGMA steering committee, and a member of the Florida International University's CHEPREO external advisory board. Brown is a founding member of GLIF and is co-chair of the GLIF Research & Applications (RAP) working group along with Larry Smarr.



Thomas A. DeFanti, Ph.D., at the University of Illinois at Chicago, is a director of the Electronic Visualization Laboratory (EVL). At the University of California, San Diego, DeFanti is a research scientist at the California Institute for Telecommunications and Information Technology (Calit2). DeFanti received a Ph.D. in Computer and Information Science from The Ohio State University in 1973. He shares recognition along with EVL director Daniel J. Sandin for conceiving the CAVE virtual reality theater in 1991. Striving for a more than 15 years now to connect high-resolution visualization and virtual reality devices over long distances, DeFanti has collaborated with Maxine Brown of UIC to lead state, national and international teams to build the most advanced production-quality networks available to academics.



Rollin Feld is Director of Operations at Skywalker Sound.



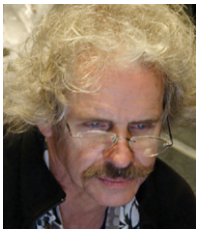
Jacob Balser is the IT Operations Supervisor and Senior Network Engineer at Skywalker Sound. As a thirteen-year IT professional, he is responsible for designing and implementing network solutions for rich media and film content. He earned his BSA in eBusiness from the University of Phoenix in 2005.



Steve Morris is Chief Engineer at Skywalker Sound.



Trevor Henthorn attended the University of California at San Diego, where he majored in Electrical and Computer Engineering and Music. He worked at the Center for Music Experiment and later as the lead systems and network administrator for the UCSD Electrical and Computer Engineering Department. He has worked with CalIT2 to help develop the "STRoke-DOC" telemedical system and worked with faculty in the UCSD Psychology Department to publish articles on psychoacoustics in Journal of Acoustical Society of America and Music Perception. Mr. Henthorn now resides in the UCSD Department of Music.



Greg Dawe is Principal Development Engineer at the University of California, San Diego division of Calit2, responsible for the design and installation of its high-performance audio and visualization technologies; his unique combination of talents includes mastery of electronics, optics, video technology, material fabrication, computers and software. Previously Dawe was Manager of System Services at the Electronic Visualization Laboratory (EVL) at the University of Illinois at Chicago, where he was principal design engineer of all of EVL's virtual-reality systems, notably the CAVE and its derivatives. Dawe has also served as one of the key technical designers for all iGrid workshops.



Peter Otto holds a Master's Degree in composition from the California Institute of the Arts. After several years serving on the faculty at Cal Arts, Otto was appointed Director of Musical Technology at the research institute Tempo Reale in Florence, Italy, where he resided for 3 years. Upon retuning to the US, he served for 4 years as Associate Professor (Composition) and Director of the Computer Music Studios at SUNY Buffalo, where he was a principal investigator for research projects including "MixNet", a comprehensive networked digital audio production system, and SUNY's Music Library Information Network project. Since 1994 Otto has served on the faculty and as Director of Music Technology at University of California San Diego's Department of Music.



Larry Smarr is the founding Director of Calit2, a partnership of the Universities of California at San Diego (UCSD) and Irvine (UCI), and is the Harry E. Gruber Professor of Computer Science and Information Technologies in the UCSD Department of Computer Science and Engineering. He was founding Director of the National Center for Supercomputing Applications (NCSA) at the University of Illinois at Urbana-Champaign in 1985, and founding Director of the National Computational Science Alliance in 1997. Smarr is principal investigator of the National Science Foundation (NSF)-supported OptIPuter initiative, co-chair of the GLIF Research & Applications (RAP) working group, and was the co-host of iGrid 2005 with Ramesh Rao, the Director of the UCSD Division of Calit2.