JANUARY/FEBRUARY | 2024
Vol. 133 No. 1

MI
MOTION IMAGING JOURNAL

INSIDE:

- Green Video Compression for Metaverse: Lessons Learned from VP9 and HEVC
- Returning Greatness to Film Grain: Introducing AV1-Compatible Film Grain Modeling for Existing HEVC-Based Video Codecs
- Profiling the ST 2110 Network Traffic for Load Testing with Open-Source Tools
- And More....

Video Compression Efficiency

Significant Advancements Enhance Consumer Experiences

COVERING EMERGING TECHNOLOGIES FOR THE GLOBAL MEDIA COMMUNITY

SMPTE

New Solutions for Desktop I/O, openGear and Color

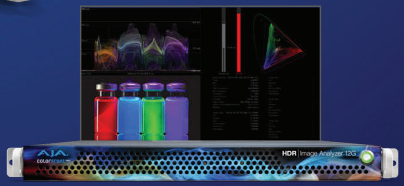
New KONA® X

- Capture/Playback 4K/UltraHD/2K/HD/SD up to 60p, as low as sub-frame latency
- Dual bi-directional 12G-SDI connections with 16-channel embedded audio
- Dual HDMI 2.0, 1x for input and 1x for output
- VESA resolution support on HDMI I/O
- YCbCr 4:2:2 10-bit and RGB 4:4:4 12-bit
- Support for OEM applications with AJA SDK and In-Firmware Microcontroller
- Supports graphics with alpha channel
- HLG and HDR10 support via HDMI
- Includes support for Apple silicon
- Optional KONA Xpand card provides expanded audio and control connectivity

New openGear® Converters

Four new openGear Converters:

- **OG-Hi5-12G**
  - 12G-SDI to HDMI 2.0
- **OG-HA5-12G**
  - HDMI 2.0 to 12G-SDI
- **OG-12G-AM**
  - 12G-SDI, AES/EBU embed/disembed
- **OG-12G-AMA**
  - 12G-SDI, analog audio embed/disembed
- All models include DashBoard support
- OG-Hi5-12G & OG-HA5-12G include an SFP port with available Fiber LC and ST options
- Compatible with openGear rackframes
- Rear I/O card included



HDR Image Analyzer 12G v3.0 Update

Free update for HDR Image Analyzer 12G offering powerful new features:

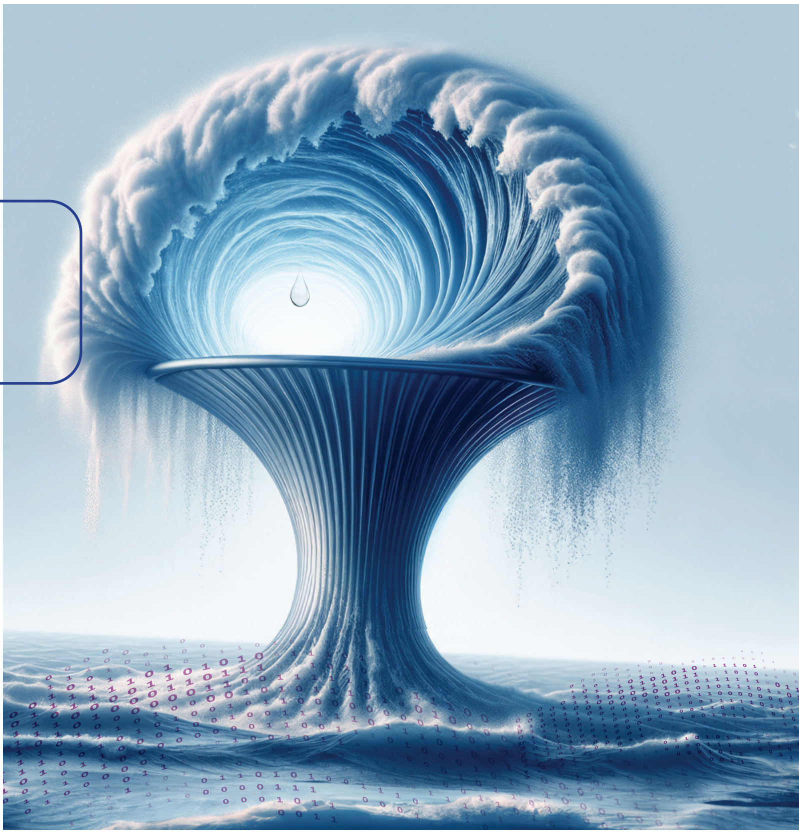
- Multi-channel signal analysis
  - Up to four channels of 4K/UltraHD
- NDI® input
  - Mix and match analysis of NDI and SDI
- Dolby Vision dynamic metadata analysis
  - Display of SDI embedded DV metadata
- ARRI LogC4 support
  - Scene referred camera Log analysis
- Additional 8K video format support

KONA X image text rotated for clarity

www.aja.com/whatsnew

MI
MOTION IMAGING JOURNAL

VOLUME 133 | NUMBER 1 | JANUARY/FEBRUARY 2024



INTRODUCTION

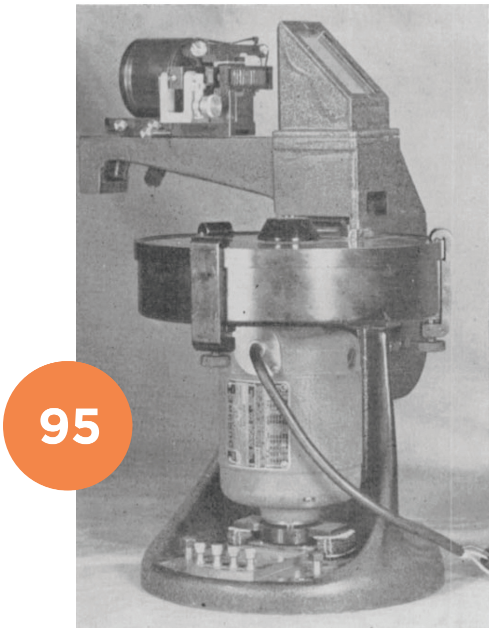
TECHNICAL PAPERS

APPLICATIONS/PRACTICES

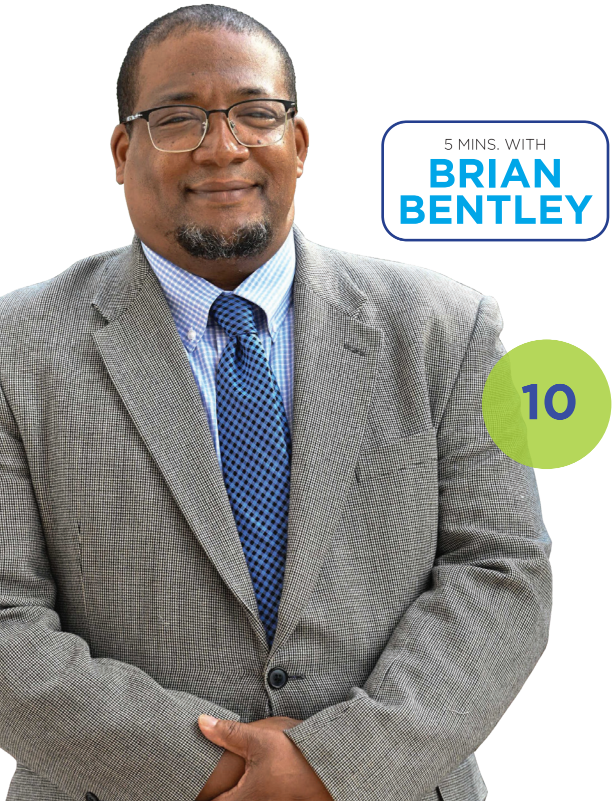
OFF-TOPIC: APPLICATIONS/PRACTICES

OFF-TOPIC: OPINION

1545-0279/24©2024SMPTE



DEPARTMENTS



The Society of Motion Picture and Television Engineers, Inc., 445 Hamilton Ave., Suite 601, White Plains, NY 10601.

EXECUTIVE
Mike Zink
Education Vice President

Jaclyn Pytlarz
Chair/Board of Editors

John Belton
Thomas Edwards
James E. O'Neal
Glen Pensinger
Associate Editors

David Grindle
Executive Director

JOURNAL STAFF
Maja Davidovic
Director of Education
education@smpte.org

Dianne Purrier
Managing Editor, Digital and Print Publications
journal@smpte.org

Anthony Catalano
Executive Account Manager, Advertising Sales & Sponsorships
sales@smpte.org

Jeffrey Victor
Advertising Sales North America
jvictor@smpte.org

SMPTE Motion Imaging Journal (ISSN: 1545-0279) published 10 times per year by the Society of Motion Picture and Television Engineers, Inc., 445 Hamilton Ave., Suite 601, White Plains, NY 10601. Periodi-cals postage paid at WhitePlains, NY, and at additional mailing office. Copyright © 2024 by the Society of Motion Picture and Television Engineers, Inc. No part of this issue may be reproduced without written permission of the publisher. Statements and opinions expressed in articles or editorials in the SMPTE Motion Imaging Journal are expressions of contributors and do not necessarily represent the policies or opinions of the Society. A complete list of SMPTE membership types and benefits is available at www.smpte.org. Advertisements appearing in this publication are the sole responsibility of the advertiser. Subscription rates: SMPTE Members, $35 USD, included in membership dues; print subscription: $1,000 (shipping included); SMPTE owns the copyrights to all published material, unless otherwise stated. *SMPTE Motion Imaging Journal* is registered with Canada Post: CPM# 41003019.

POSTMASTER: Send change of address to *SMPTE Motion Imaging Journal*, 445 Hamilton Ave., Suite 601, White Plains, NY 10601. All rights reserved. PRINTED IN THE USA

DOI: 10.5594/JMI.2024.KSPN8355
Date of publication: 24 January 2024

SMPTE OFFICERS

OFFICERS

PRESIDENT, 2023–24
Renard T. Jenkins

PAST PRESIDENT, 2023–24
Hans Hoffmann
EBU

EXECUTIVE VICE PRESIDENT, 2023–24
Richard Welsh
Deluxe Media Inc.

STANDARDS VICE PRESIDENT, 2024–25
Sally Hattori
Walt Disney Studios

EDUCATION VICE PRESIDENT, 2023–24
Michael Zink
Warner Bros. Discovery

MEMBERSHIP VICE PRESIDENT, 2024–25
Rosemarie Lockwood
Belden

FINANCE VICE PRESIDENT, 2024–25
Paul Stechly
Applied Electronics

SECRETARY/TREASURER, 2023–24
Lisa Hobbs
MediaKind

EXECUTIVE DIRECTOR
David Grindle
SMPTE Headquarters

DIRECTORS

STANDARDS
TBD

EDUCATION
TBD

MEMBERSHIP
TBD

GOVERNORS

GOVERNORS-AT-LARGE
TBD

ASIA/PACIFIC REGION
Michael Day, 2024–2025
Telstra Corporation

Tony Ngai, 2023–24
Society of Motion Imaging Ltd.

CANADIAN REGION
Tony Meerakker, 2024–2025
Meer Tech System

François Bourdua, 2023–24
VS-Tek, Strategies & Technologies

USA - CENTRAL REGION
William T. Hayes, 2023–24
Iowa Public Television

USA - EASTERN REGION
David Long, 2024–25
MAGICSpell Studios, RIT

Dover Mundt, 2023–24
Riedel Communications

USA - HOLLYWOOD REGION
Eric Gsell, 2023–24
Dolby Laboratories

Marty Meyer, 2023–24
Gravity Media

Kylee Peña, 2024–25
Adobe

EMEA, CENTRAL & SOUTH AMERICA REGION
Fernando Bittencourt, 2023–24
FB Consultant

Dagmar Driesnack, 2024–25
Rohde & Schwarz GmbH & Co. KG

USA - NEW YORK REGION
Thomas Mauro, 2023–24
Media Consultant

USA - SOUTHERN REGION
Frank Torbert, 2024–25
WKMG-TV

U.K. REGION
Chris John, 2023–24
Sky UK

USA - WESTERN REGION
Jeffrey F. Way, 2023–24
Quantum

Cassidy Lee Phillips, 2024–25
swXtch.io.

SECTION CHAIRS

ATLANTA
Jack Linder
Linder Consulting

AUSTRALIA
Paul Whybrow
Bodyboard Immersive Experiences

CHICAGO
Jim Skupien
Riedel

DETROIT
R. Thomas Bray
University of Michigan

FLORIDA
Shawn Maynard
Florical Systems

GERMANY
No Section Chair

HOLLYWOOD
Marc Zorn
Marvel Studios

HONG KONG
Terence Yiu
Phoenix Satellite Television Co. Ltd.

INDIA
Ujwal Nirgudkar

ITALY
Alfredo Bartelletti
BLT Italia srl

MONTREAL/QUEBEC
David Beaulieu
CBC / Radio-Canada

NASHVILLE
No Section Chair

NEW ENGLAND
Martin P. Feldman
Unique Scientific, Inc.

NEW YORK
Ed DeLauter
Paramount

NORDIC
No Section Chair

OHIO
John R. Owen
Q Communications, LLC

PHILADELPHIA
Mark Mullen
Millersville University

PACIFIC/NORTHWEST
No Section Chair

POLAND
Kamil Rutkowski
Black Photon Sp. z o.o.

ROCKY MOUNTAIN
Kerry Booth
Signiant

RUSSIA
Suspended

SACRAMENTO
Paul Turner
AJA Video

SAN FRANCISCO
Kent Terry

TEXAS
Curtis Kirk,
PlayMate Enterprise LLC

TORONTO
Lawrence St-Onge
LSO Consulting

UNITED KINGDOM
John Ellerton
BT Media & Broadcast Dolby

WASHINGTON, DC
Maciej Ochman
Corporation for Public Broadcasting

STUDENT CHAPTERS BY SECTION

ATLANTA
Clark Atlanta University
Atlanta, GA

AUSTRALIA
Sydney Student Section
Sydney, Australia

DALLAS
Baylor University
Waco, TX

CHICAGO
Purdue University
West Lafayette, IN

FLORIDA
Full Sail University
Winterpark, FL

GERMANY
Universities of Applied Sciences Germany (HdM / HSHL / HSRM)

- Hochschule der Medien | Jan Fröhlich
- Hamm-Lippstadt University of Applied Sciences | Stefan Albertz
- RheinMain University of Applied Sciences | Wolfgang Ruppel

HOLLYWOOD
Los Angeles City College
Los Angeles, CA

Media Arts Collaborative Charter School
Albuquerque, NM

Pasadena City College
Pasadena, CA

USC University of Southern California
Los Angeles, CA

University of Nevada, Las Vegas
Las Vegas, NV

HONG KONG
Hong Kong Design Institute (HKDI)
Hong Kong

Hong Kong Institute of Vocational Education (IVE)
New Territories, Hong Kong

INDIA
Whistling Woods International
Mumbai, India

NEW ENGLAND
Fitchburg State University
Fitchburg, MA

Stonehill College
Easton, MA

NEW YORK
Barnard College
New York, NY

Borough of Manhattan Community College
New York, NY

Bronx Community College
Bronx, NY

Kingsborough Community College
Brooklyn, NY

New York City College of Technology
Brooklyn, NY

Rowan University
Glassboro, NJ

William Patterson University
Wayne, NJ

PHILADELPHIA
Millersville University
Millersville, PA

ROCHESTER
Rochester Institute of Technology
Rochester, NY

SAN FRANCISCO
De Anza College
Cupertino, CA

City College of San Francisco & San Francisco State University
San Francisco, CA

TEXAS
University of North Texas

UNITED KINGDOM
Birmingham City University
Birmingham, U.K.

Ravensbourne
London, U.K.

Southampton Solent University
Southampton, Hampshire, U.K.

University of Salford
Manchester, U.K.

University of Surrey
Guildford, U.K.

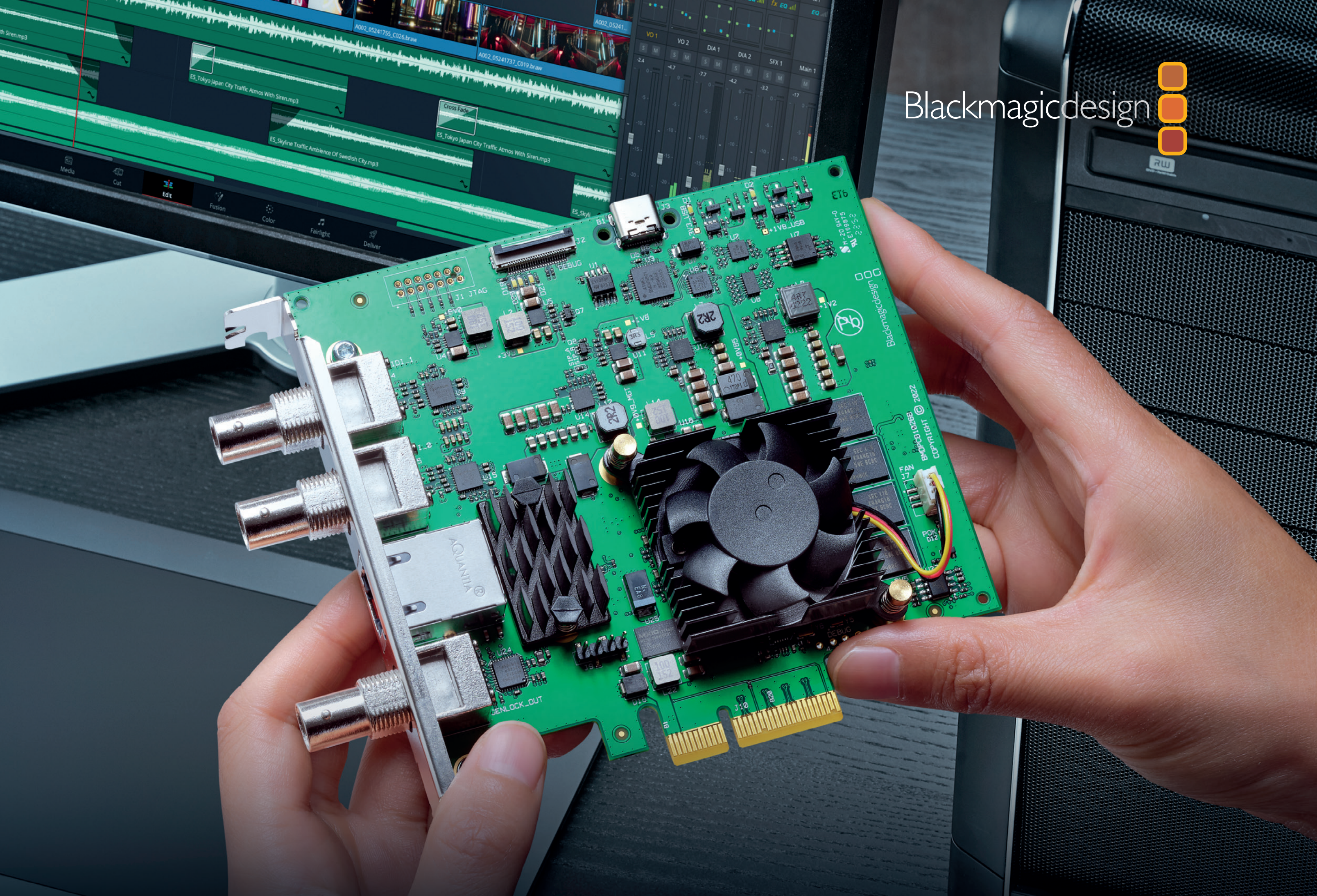
WASHINGTON, DC
George Mason University
Fairfax, VA

Loyola University
Baltimore, MD

Montgomery College
Rockville, MD

BOARD OF EDITORS

1545-0279/24©2024SMPTE



The secure way to capture and play video direct to 2110 IP systems!

DeckLink IP cards are the easiest way to capture and play back video directly into 2110 IP based broadcast systems! DeckLink IP cards support multiple video channels plus each channel can capture and play back at the same time. You can build racks of servers generating broadcast graphics, virtual sets, or GPU based AI image processing directly integrated into 2110 IP broadcast infrastructure.

10G Ethernet for Multiple SMPTE 2110 IP Channels

DeckLink IP cards conform to the SMPTE ST2110 standard for IP video, which specifies the transport, synchronization and description of video, audio and ancillary data over managed IP networks for broadcast. DeckLink IP supports SMPTE-2110-20 uncompressed video, SMPTE-2110-21 traffic shaping/timing, SMPTE-2110-30 audio and SMPTE-2110-40 for ancillary data.

High Speed 4 Lane PCIe Connection

With a high speed 4 lane PCI Express interface, DeckLink IP cards feature a PCIe connection to the host computer that's fast enough to handle multiple HD video channels, as well as simultaneous capture and playback on each of the channels. With PCI Express, you get very low latency combined with highly accurate time synchronization to the software that's capturing or playing video.

Supports All SD and HD Formats up to 1080p60!

DeckLink IP connections are multi rate so they support all SD and HD formats. In SD, DeckLink IP supports both 525i59.94 and 625i50 standards. In HD, DeckLink IP cards support all 720p standards up to 720p60, all 1080 interlaced standards up to 1080i60 and all 1080p standards up to 1080p60. There's no complex change over as the cards can switch instantly to a new video standard.

Develop Custom Solutions with the DeckLink SDK

Available free with all DeckLink cards is an advanced developer SDK for Mac, Windows and Linux that you can download free of charge! The DeckLink SDK provides low level control of hardware when you need flexibility, as well as high level interfaces which are a lot easier to use when you need to perform more common tasks.

DeckLink IP HD	$345
DeckLink IP HD Optical	$345
DeckLink IP/SDI HD	$395

Learn more at www.blackmagicdesign.com

CORPORATE MEMBERS

DIAMOND LEVEL

Amazon AWS
Blackmagic Design, Inc.
CBS, Inc.
Deluxe
Disney/ABC/ESPN
Dolby Laboratories
Fox Corporation
Google
Paramount Pictures
Ross Video
Sony Electronics, Inc.
Telstra Corporation
Warner Bros. Discovery

PREMIUM LEVEL

Academy of Motion Picture Arts & Sciences
Apple, Inc.
ATOS IT Services UK
AVID Technology, Inc.
Bloomberg
British Telecommunications, PLC
Imagine Communications
Microsoft Corp.
NBC Universal
Rohde & Schwarz, Inc.

ADVANCED LEVEL

AJA Video Systems Inc.
AMD
Belden, Inc.
Bridge Technologies
Canon, Inc.
Dell
Densitron Technologies
European Broadcasting Union
Fuse Technical Group
Huawei
Interdigital Communications
Library of Congress
MediaKind (formerly Ericcson Media Solutions)
Mo-Sys Engineering
NEP Group
Novastar
NTC, A Deloitte business
Panasonic Corporation
Qube Cinema
Red Digital Camera
Roe Visual Co, Ltd.
Seagate Technology
Signiant
Sky U.K.
Streamland Media/Picture Shop
Sudwestrundfunk / ARD
Texas Instruments
The Studio - B&H
Xperi

ESSENTIAL LEVEL

4Wall Entertainment
AOTO
Applied Electronics Ltd.
Arqiva Ltd.
ARRI, Inc.
Astrodesign Inc.
Brompton Technology
Canare
Carl Zeiss AG
CBC Radio Canada
Chambre des Communes
Channel 4 Television
Cisco
Cooke Optics
Creamsource
Dalet Digital Media Systems
Digital TV Group (DTG)
Disguise
Disney Streaming Services
Diversified
Ericsson
Evertz
EVS/Broadcast Equip
Extreme Reach
Fraunhofer
Grass Valley, Inc.
ICVR
IMG Media
Intel Corporation
Koninklijke Philips NV
Ledyard/Planar
Matrox Electronic Systems, Ltd.
Media Links Co., Ltd.
MediaSilo
Megapixel VR
Meinberg-Funkuhren GmbH & Co.
Motion Picture Solutions
MLB Advanced Media
National Association of Theater Owners
NEC Corporation
Net Insight
Nevion
NHK (Japan Broadcasting Corp.)
Nvidia
Pebble Beach Systems
Perforce Systems
Phabrix Ltd.
Pixelogic
pixit media
Pixotope
Portrait Displays
Quasar Science
Qube Cinema
Riedel Communications
Rosco Laboratories
Schweizer Radio und Fernsehen
Sencore, Inc.
Synamedia
Synaptics, Inc.
Tag VS
Telestream, Inc.
Universal Pictures
V-Nova
Vu¯
XR Studios
Yleisradio Oy
Zixi

SMALL BUSINESS LEVEL

Adder Technology
Adeas, B.V.
Amphenol RF
Appear TV AS
Arista Networks
ATEME
Australian Institute of Aboriginal & Torre Strait Islander Studies (aiatsis)
Aveco
Barco
BBC Future Media
Boland Communications
BLT Italia srl
Broadstream Solutions
Castlabs GmbH
Chesapeake Systems
CineCert
Cobalt Digital
CST (Comission Superiere Technique de l'image et du son)
DekTec America
Deltacast.tv
Disk Archive Corporation Limited
DSC Laboratories
Eikon Group Co.
Eluv.io
Flanders Scientific
Fujifilm Inc.
GDC Technology
Glassbox Technologies
IHSE USA LLC
Imagica Entertainment Media Services, Inc.
Innovative Production Services
InSync Technology Ltd.
Intelligent Wave Inc.
Internet Initiative Japan
IntoPIX
Kino Flo, Inc.
LAWO
Leader Instruments Corp.
LG Electronics
Light Field Lab, Inc.
Lynx Technik AG
Macnica Technology
Marquise Technologies
Media Tek Inc.
Merrill Weiss Group, LLC
Metaglue
Mole-Richardson Co.
MTI Film
Netgear AV
The Nielsen Company (US), LLC
NTT Network Innovation Labs
Original Syndicate
Panamorph
Plus 24
Port 9 Labs
Qvest Gmbh
Raysync
Seiko Epson Corp.
Showfer Media LLC
Soliton Systems
SRI International Sarnoff
Starfish Technologies
Sutro Tower, Inc.
Tajimi Electronics Co., Ltd.
Tamura Corporation
Techex Ltd.
Tedial
Teledyne Lecroy PSG
Telos Alliance
The Family Collective, LLC
Tokyo Broadcasting System Television, Inc.
TSL Professional Products Ltd.
Utah Scientific
Video Clarity, Inc.
Visionular

INDUSTRY PARTNERS

IBC
ITU-R
NAB
NISO

1545-0279/24©2024SMPTE

FROM THE TOP

SMPTE Executive Director's Message

A New Year

and Exciting Enhancements to the *Motion Imaging Journal* and SMPTE Membership

DAVID GRINDLE

Starting the new year with an exciting new look for the *Motion Imaging Journal* is a great way to share the energy and ideas of SMPTE members. The new look of this journal comes from feedback and input from our editorial board and journal staff members. The journal has had many "refreshers" over its history. This is just one in a series of modifications. What remains constant is the peer-reviewed content.

Peer-review has long been used in science and technology to ensure that accurate and honest information is shared. Applying this process to our journal and technical papers gives SMPTE a global reputation for solid science. I am thankful to the peer reviewers who make this process possible and the authors who put in the time and effort and are willing to have their work scrutinized by their peers. Our goal is to keep increasing the number of authors that submit work for peer review to get the latest research from related fields and those from our historical pool of media scientists and technologists. Outreach to computer science, physics, and engineering professionals to submit their work is key to maintaining a strong and healthy journal.

Our members also noted that they would like use cases and other types of articles. This, too, is important. Many of our members work in the field of applied media science daily. Those stories of success, discovery, and sometimes failure are equally important. They require editors but not necessarily a scientific peer review. It makes them no less important to our members. We are excited to expand the journal to include these articles.

Including these adds value for our readers globally while opening the door for new authors to share their experiences with the SMPTE membership.

Adding these updates suggested by our members keeps the members' voices centered in SMPTE. Paired with the inclusion of the full Standards Library Access and Self-Paced Learning in the cost of membership for each person, SMPTE is reinventing its view of member benefits. These reflect the reality of what people look for in membership organizations.

Our Standards community continues its work on revisions and updates to Standards and discusses the impact of AI and the move to open-source environments. Both will impact the traditional systems for creating Standards but will not negate the need for Standards. This remains part of SMPTE's core.

In addition to Standards, our members seek support in learning skills and showing the success of that

1545-0279/24©2024SMPTE

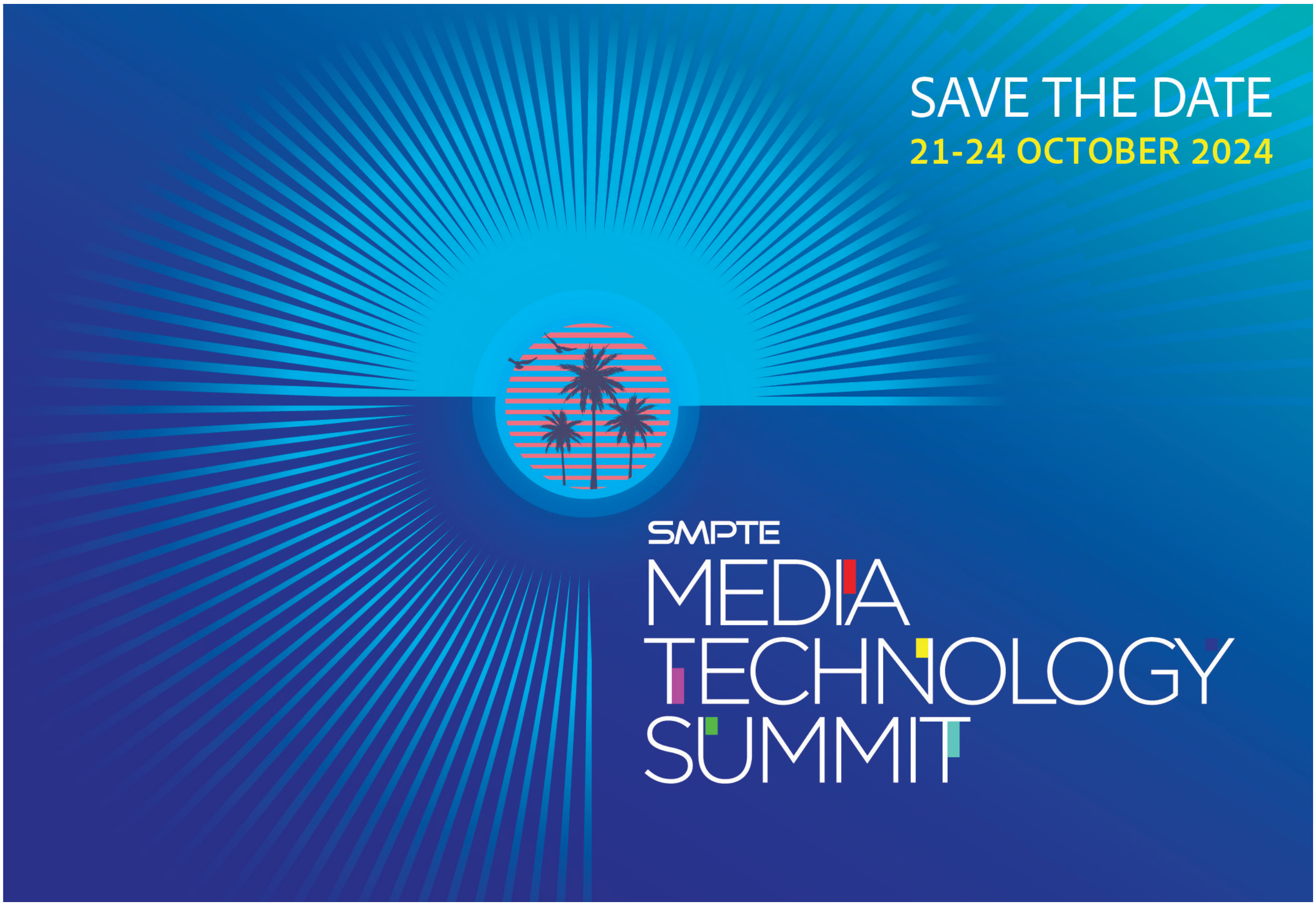
I IMPLORE YOU **TO BE ACTIVE AND ENGAGED** IN YOUR CHOSEN SECTION, CHAPTER, OR GLOBAL EFFORT.

learning. We are working to provide those kinds of certificates that can be shared and acknowledged by employers. This is yet another member request that we are working to address to provide the value that members seek in being part of SMPTE.

With all this said, I implore you to be active and engaged in your chosen Section, chapter, or global effort. We need your constructive ideas and input. As members, you can best articulate what brings value to your work environment. As a global organization, we must acknowledge that what one part of our membership needs, another may find less valuable. But that doesn't mean there is no value. We may be like the American Automobile Association, providing a myriad of benefits used by some of our members and not by others, yet offering value to all.

I hope you enjoy the refreshed version of the *Motion Imagining Journal* and look forward to continuing to bring value to your SMPTE membership.

DOI: 10.5594/JMI.2024.FROO4424
Date of publication: 24 January 2024



1545-0279/24©2024SMPTE

SMPTE Spotlight:

Brian Bentley, EdD

BY RUSSELL POOLE

Some of SMPTE's most influential members are faculty at highly esteemed colleges and universities. One of these members is Brian Bentley, EdD, Associate Dean for the School of Arts & Sciences and Assistant Professor in the Department of Mass Media Arts at Clark Atlanta University. Bentley has had an incredible career and is responsible for creating the first Historically Black Colleges and Universities (HBCU) SMPTE student chapter.

Bentley began his career as a student, earning a BA in broadcasting from Southern University and a Certificate in game design from Full Sail University. He has an incred-

CURRENT POSITION:
Associate Dean for the School of Arts & Sciences and Assistant Professor in the Department of Mass Media Arts at Clark Atlanta University

PROFESSIONAL ORGANIZATIONS:
SMPTE; Broadcast Education Association (BEA); National Association of Black Journalists (NABJ); Association for Education in Journalism and Mass Communication (AEJMC); National Academy of Television Arts & Sciences (NATAS); Southeast Chapter, International Game Developers Association (Atlanta Chapter); National Association of Mathematicians

DEGREES:
MA, MBA, MFA

ible number of degrees, including an MA in mass communications from Southern University, an MBA from Purdue University Global, and an MFA in creative writing from Full Sail University. He received a doctorate in instructional leadership from Argosy University. Bentley is also now pursuing a PhD in strategic media at Liberty University.

“I would say that my main influences during my career were my parents,” said Bentley when asked about those who influenced him. “They always instilled in me the value of hard work and taught me always to be professional. My parents are always there for encouragement and inspiration.”

While earning his degrees and after receiving them, Bentley had a rather illustrious career in film and television, working as a television news producer while sharing time in production. Furthermore, he has worked in the post-production industry, spending some time as a film colorist. His career in education spans more than 20 years. Bentley's roles have included administrative positions such as Associate Dean for Arts & Sciences, Campus Director of Academic Operations, Associate Dean of Academic Affairs, and Department Chair for Digital Filmmaking/ Video Production & Visual Effects and Motion Graphics. He has also served as co-director of broadcast training for radio, television, film and full-time faculty. Positions like these gave him the gravitas and authority to start the first SMPTE student chapter at an HBCU. “The first HBCU SMPTE chapter came to be from the inquiry, ‘What do you all think about a student SMPTE chapter at Clark Atlanta University?’” Bentley spoke on the student chapter’s founding. “The idea was well received because it would be an opportunity to host industry meetings and student workshops. Having the first HBCU student chapter at Clark Atlanta University is awesome and exciting!”

Bentley’s work at Clark Atlanta University (CAU) and with SMPTE is transforming the entire industry. He is one of the leading forces behind the inaugural “Power of Color Symposium,” a SMPTE event dedicated to the science of representation in media. Bentley and SMPTE came together to create an amazing event, which takes place at CAU in February.

“Through this collaboration with SMPTE, I hope that we can foster opportunities for both students and faculty in the area of professional development,” said Bentley when asked about his hopes for the future. “In addition, I would like to leave a lasting impact by always being a team player and working to prepare our students to be competitive in a global job market with industry-level skill sets.”

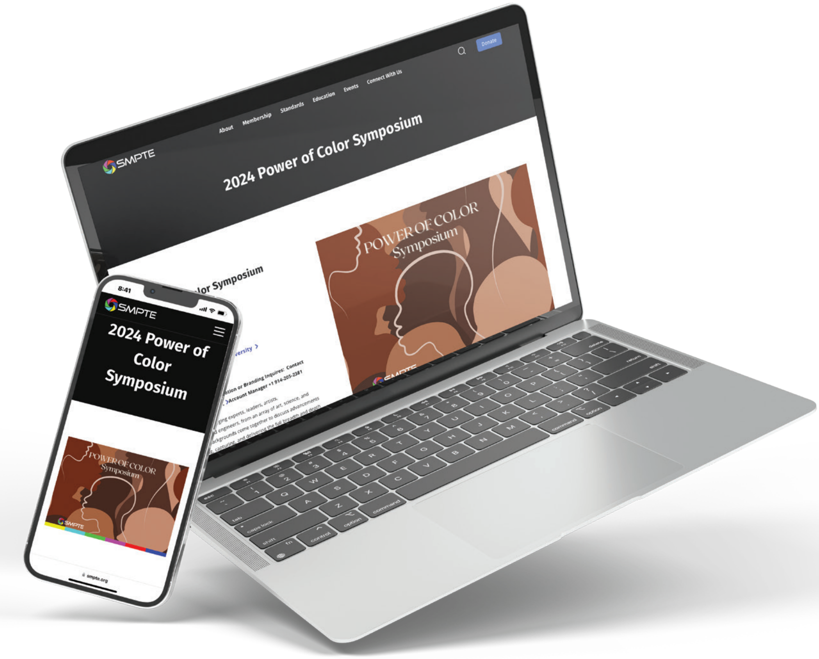
> “I WOULD LIKE TO LEAVE A LASTING IMPACT BY ALWAYS BEING A TEAM PLAYER AND WORKING TO PREPARE OUR STUDENTS TO BE COMPETITIVE IN A GLOBAL JOB MARKET WITH INDUSTRY-LEVEL SKILL SETS.”

In addition to SMPTE, Bentley is a member of several organizations, including the Broadcast Education Association (BEA), National Association of Black Journalists (NABJ), Association for Education in Journalism and Mass Communication (AEJMC), National Academy of Television Arts & Sciences (NATAS) Southeast Chapter, International Game Developers Association (Atlanta Chapter), and National Association of Mathematicians. He also enjoys watching sports and movies, hiking, and learning new technologies.

DOI: 10.5594/JMI.2024.PRBC4004
Date of publication: 24 January 2024

NEW AND UPCOMING:

SMPTE to Host First-of-a-Kind Power of Color Symposium


SMPTE
About Membership Standards Education Events Connect With Us
Donate
2024 Power of Color Symposium
POWER OF COLOR Symposium
8:41
SMPTE
2024 Power of Color Symposium
POWER OF COLOR Symposium
smpte.org

SMPTE will host its inaugural Power of Color Symposium at Clark Atlanta University in Atlanta, Georgia, on 6-7 February 2024, the beginning of Black History Month. With its unique focus on people of color, the Power of Color Symposium will bring together established and emerging experts, artists, technicians, engineers, and other leaders from the film, television, animation, and gaming industries to discuss advances in designing, capturing, and delivering the full breadth and depth of tone, texture, and vibrancy seen in global humanity.

The two-day event will feature topics such as the importance of representation; cross-cultural communication as it pertains to viewing skin tones; the art of representation; scientific and technical tools for skin tones, color tones, hair textures, hues, and makeup; SMPTE Rapid Industry Solutions (RiS) Color Management and color workflows from camera lens to screen; and the future of representation within the field.

"SMPTE has long been known for its color bars, and now the Society is exploring the power of color to raise the bar for accurate, equitable, and inclusive visual representation across media and entertainment," said Michele Wright, director of business development and outreach at SMPTE and chair of the Power of Color Symposium. "There is code in color — skin tones and hues, eye color, hair texture — and

REGISTER TO ATTEND

Power of Color Symposium
Atlanta, Georgia
February 6-7, 2024

1545-0279/24©2024SMPTE

value in culture, both of which can help to convey the essence of a person or character and what they represent. Connecting people and diverse communities to explore the creative, technical, and even cultural and sociological aspects of effective color representation, this inaugural symposium brings further depth and dimension to the Society's annual conferences and events, in turn speaking to the needs and interests of media professionals around the globe."

Through in-person and online panel discussions, sessions, and hands-on workshops, SMPTE's Power of Color Symposium will amplify the voices, insights, and foresight of industry leaders and change agents committed to exploring and expanding topics, conversations, and techniques that advance the power of effective color representation throughout the media and entertainment industry and across a multitude of ethnicities and cultures.

"For anyone dedicated to the craft of storytelling – the stories of people and humanity – the Power of Color Symposium is an important opportunity to exchange ideas, acquire knowledge, build critical skills, and form valuable connections," said Brian Bentley, associate dean, Arts & Sciences, at Clark Atlanta University, which recently became the first HBCU (Historically Black Colleges and Universities) with a SMPTE Student Chapter. "We're proud to host this remarkable event and to support SMPTE, and creatives and technologists everywhere, in raising the bar for color and representation in moving pictures."

"While we've seen progress worldwide with respect to diversity and representation in media, it's also clear that we still have a long way to go and must build on those successes – learn from experienced professionals across lighting, makeup, animation, visual effects, and other fields to understand what matters, and why," said Renard T. Jenkins, SMPTE president. "The Power of Color Symposium offers access to a wealth of knowledge, delivered by a fantastic lineup of industry professionals, and participants will come away with a better understanding of both the technical and creative considerations associated with achieving accurate visual representation that resonates with viewers. It's going to be a dynamic two days of learning, dialog, and connection."

DOI: 10.5594/JMI.2024.VLUN9330
Date of publication: 24 January 2024

The 2023 MTS Showcases New Features, Technologies, and Faces

BY RUSSELL POOLE

The media technology industry never stops changing, and SMPTE is evolving with it. That's why this year's Media Technology Summit was full of new features and presenters, showcasing the power and influence of new technology and the innovators who create it.

Over the course of four days (16-19 October) attendees experienced detailed technical sessions, illuminating special sessions, and special networking events that provided the Summit with a spirit of professional comfort. Attendees raved about this year's Summit, stating it had a more relaxed atmosphere that allowed them to enjoy everything at their own pace. This made for a fun, educational Summit, which is just what the media industry needed this year.

"The Energy at the 2023 MTS was exciting," said SMPTE Executive Director, David Grindle. "From Hanno Basse's keynote to the end of the Gala, people were networking and learning. We tried new things and many of them worked. I look forward to seeing how we can build upon the lessons learned to continue making this event impactful for the attendees."

Here are the highlights of the sessions and events, and new features introduced at this year's conference.

Monday, 16 October – Day 1

The Summit kicked-off with a powerful introduction from SMPTE President Renard T. Jenkins, Summit Co-Chairs Iris Wu and Zandra Clark, and David Grindle. "This was an extraordinary moment in the Society of Motion Pictures and Television Engineers history at this year 2023 Media Technology Summit," said Summit Co-Chair, Zandra Clark. "Post covid, there was a career professional surge that occurred as Life timers, Boomers, Gen X's, Millennials, and Gen Z converged to share a common objective: To learn more about the future of media and content innovation. And this Summit delivered!"

Following the introduction, Hanno Basse, CTO of Digital Domain, gave a keynote address on Generative AI and

1545-0279/24©2024SMPTE

SMPTE MEDIA TECHNOLOGY SUMMIT
Generative AI and Machine Learning for the Creation of Digital Humans and Other Visual Effects Assets
Hanno Basse, CTO, Digital Domain
Hollywood California #MTS2023
on here with all the new technology that's emergin

DAY 1: Hanno Basse provided the keynote address to kickoff the 2023 MTS.

Machine Learning for the Creation of Digital Humans and Other Visual Effects Assets. The talk focused on the benefits of AI in visual effects (VFX), including digital humans. Most notably, he stressed that Generative AI will never replace human beings as actors. "The human audience wants to connect with human beings," said Basse. "Every example has the performance of a real actor underlying it. That's how you get the real emotion and performance. I don't see that changing."

Monday's technical sessions began immediately following the keynote. The day's topics ranged from streaming technology to compression. Session chairs included Ievgen Kostiukevych of the European Broadcasting Union, Dagmar Driesnack of Rohde & Schwarz, Jason Thibeault of the Streaming Video Technology Alliance, Marc Zorn of Marvel Studios, Juan Reyes of Cloud Security and Assurance, and Jaclyn Pytlarz of Dolby Laboratories.

Some of these sessions included "Cameras and LED Walls—A Challenging Relationship," in which Klaus Weber of Grass Valley discussed interoperability issues in virtual production and shared several solutions to said issues. Eric Diehl of Sony Pictures shared a more efficient and comprehensive way to release films in "Atheneum—A Blockchain to Manage Theatrical Releases," and in "Breaking the Fourth Wall Through Extended Reality," Jason Kao and Karen Kao from Inland Norway University of Applied Sciences and Glasgow Caledonian University London, respectively, discussed versatile storytelling by using extended reality.

The annual Section Leadership meeting concluded Day 1. SMPTE Section leaders, whether in attendance at the Summit or remotely, received an update on the current state of the organization, as well as its plans to serve the industry and its members. In-person attendees enjoyed an ice cream Sunday bar by the pool of the Loews Hollywood Hotel.

Tuesday, 17 October – Day 2

The second day of the Summit began with the unveiling of two new, huge features: The Solutions Hub and the Emerging Tech Showcase. The Solutions Hub named to replace the exhibit hall, was full of companies showcasing solutions to the industry's biggest issues. Many of these solutions were included in the Emerging Technology Showcase, which exhibited cutting-edge technology from the most innovative companies and solutions. AI captioning from AI Media, a laser projector from Cinionic, an LED wall from QST LED, and live capture equipment from The Studio-B&H were among the companies that supplied technology for the Emerging Tech Stage.

Day 2 also marked the start of the Special Sessions, which covered a wide range of topics from SMPTE Standards to sustainability in media technology. Panel discussions, plans for SMPTE initiatives, and stories about how technology can change the world were among the topics covered in these sessions. Some of the highlights included, a "SMPTE RiS OSVP Color Management Report, which provided a comprehensive update from the color management team of SMPTE's RiS OSVP initiative. A session "From Duct Tape and Bailing Wire to SMPTE ST 2110" focused on the role students played in this year's broadcast of the Super Bowl.

In another session, "Bridging Tech and Storytelling." SMPTE's director of business development, Michele Wright, PhD, led a diverse panel in a discussion on how new technology changes how filmmakers can tell their stories. "Not only did we have two phenomenal women who co-chaired this well-rounded event, but I was also honored to lead and host a notable panel of all women of color from diverse professions, backgrounds, and experiences," said Wright. "These powerhouses astoundingly and collaboratively shared their unique experiences and perspectives about 'Bridging Tech

1545-0279/24©2024SMPTE

16 - 19 October 2023
Hollywood California #MTS2023
Lakey Campbell all the way from
New Mexico with us. A
MEDIA TECHNOLOGY SUMMIT
Welcome to the Spectrum Experience
NETWORKS
VISIBILITY

DAY 2: The unveiling of the Emerging Showcase and Solutions Hub and wide range of topics were covered on Day 2, including a special session on Bridging Tech and Storytelling.

and Storytelling,' which was just one of many phenomenal sessions led by renowned experts across the worldwide media and entertainment spectrum."

Day 2 included many technical sessions. Thomas Kernen of NVIDIA and Ievgen Kostiukevych, chaired session, with topics ranging from Cloud based media production to SMPTE 2110. Presentations included "Optimized Cloud Streaming for Ultra-Low-Latency Cloud Gaming and VR/XR Applications," in which Kevin Mockford of V-Nova discussed virtual reality capabilities on the Cloud. In "The Coronation of King Charles III," Simon Thompson of BBC Research and Development talked about the technical requirements of filming the coronation of King Charles III and Jae-Young Lee of the Electronics and Telecommunications Research Institute discussed the future of media in the Cloud in "Development of Cloud-Based Media Production Systems." The evening ended with an opening reception involving another new Summit feature: The Color Bar. This new lounge was a celebration of the 45th anniversary of the game-changing SMPTE Color Bars. Although the bar opened earlier in the evening, the Solutions Hub Opening reception began after the final technical session ended. There was abundant food and drink for all those in attendance, and this was a great way to celebrate the Summit, SMPTE, and everyone in the industry.

Wednesday, 18 October – Day 3

Day 3 marked the final day of the Solutions Hub and the Emerging Technology Showcase. There were many amazing sessions, touching on topics like display technology and SMPTE ST 2064. One of the coolest sessions on the Emerging Tech Stage was "Sustainable Virtual Production Techniques" with virtual production editor for *American Cinematographer Magazine*, Noah Radner. This session highlighted the environmental benefits of virtual production studios and was one of the best attended talks of the entire Summit.

The Special Sessions also made waves on Day 3, with topics that kept attendees on the edge of their seats. Among the sessions, "What's new in RiS-OSA?" Chris Lennon, head of OSA, presented SMPTE RiS-OSA's current work and future objectives. In another session, "Yes, Web3 is Still Relevant!" Kylee Pena of Adobe discussed how blockchain technology has a life outside of crypto-currency and NFTS, particularly in media spaces, and SMPTE Standards VP Sally Hattori discussed the future of SMPTE Standards in "SMPTE is Moving its Standards Library to HTML."

Day 3 was also the final day of technical sessions. Sessions chairs included SMPTE President Renard Jenkins, John Ferder of the SMPTE Board of Governors, Ian MacSpadden of Arizona Public Media, and Jaclyn Pytlarz of Dolby Laboratories.

Day 3 was also SMPTE Student Day! Students got special access to sessions, the Solutions Hub, and a special networking event at the end of the day. This student networking event allowed students to talk among themselves as well as established members of the media tech community. SMPTE is extremely proud of our student members,

How Current Technology Trends
Influence our Creative Choices
Marc Zorn, Marvel Studios (Chair)
Eric Weaver, USC Entertainment Technology Center
Chuck Parker, Sohonet CEO
Jen Underdahl, Marvel Studios
Greg Ciaccio, IMAX
and they're converging right? So
no
AVECO
ASTRA STUDIO

DAY 3: Panels, Emergency Tech Showcase, special sessions, and special access to students were all a part of Day 3's lineup.

1545-0279/24©2024SMPTE

DAY 4: Attendees visited virtual production studios.

and we're always glad to provide a place for them to gain more experience and connections.

Thursday, 19 October – Day 4

On Day 4, attendees were able to choose one of four off-site visits to studios (Disguise, Epic, Third Floor, and XR) for their final day, each one dealing with some aspect of virtual production.

Visitors to Disguise Studios got a firsthand look at the evolution of virtual production, as well as the role ST 2110 played in said evolution. They also showed attendees the future of LED lighting, as well as the role RGBW panels play in ICVFX.

Epic Studios guests received a deep dive into the use of SMPTE 2110 on LED walls for ICVFX workflows. The trip included a demonstration on the Epic Games LA Lab Stage and Nant Studios campus, allowing attendees to get a close look at virtual production infrastructure.

The Third Floor Visualization Studio visitors learned about the advantages of virtual production stages of varying sizes. They also examined visualization techniques, particularly in how pre-production might bring a filmmaker's vision to life.

Those who attended XR Studios had the opportunity to observe virtual production workflows and best-case scenarios. Attendees saw three examples of virtual production in action. These four trips were educational and enlightening for attendees, but the festivities didn't stop there.

The Annual SMPTE Gala, which concluded the 4-day event, celebrated 35 individuals who changed the industry forever. The evening began with a red carpet reception. The Gala ceremony was an elegant affair, complete with fine wine, dining, and revelry. The evening ended with a lively afterparty.

Conclusion

This year's SMPTE Summit had some bold new features, all of which were a massive success. The Media Technology Summit is held for the benefit of the media tech industry. It's meant to drive said industry forward, and present solutions for the biggest issues facing film and television today. We at SMPTE believe that goal was achieved and are thankful to all our sponsors and attendees for helping us make it great.

As our very own Michele Wright said, "This MTS was epic, unforgettable, and one for the history books! I look forward to continuing to work in synergy with our staff, members, students, partners, and sponsors to not only bridge the gap in tech across an array of dimensions but to help to passionately reiterate SMPTE's commitment to raising the bar (color and beyond) for not only today but many more generations and MTS' to come. The BEST is yet to come!"

DOI: 10.5594/JMI.2024.GHEN8084
Date of publication: 24 January 2024

Exploring Advances in Video Compression Efficiency

BY SEAN MCCARTHY

Compression is a foundational technology in our industry that enables content creators to deliver spectacular experiences to people around the world. It is no wonder, then, that compression is an annual theme of our Journal. This year, video compression experts tell us how existing technologies continue to improve, show us how to use the latest video coding standard, and bring us up to speed on the future of AI-powered video compression.

The overall theme of this issue is compression efficiency, but efficiency is not only about bandwidth. Energy efficiency is also critical and the subject of Natalia Molinero Mingorance's paper describing a study of motion estimation, a power-hungry aspect of video compression, to understand how energy efficiency can be improved for the high-resolution, streaming, and

THE PAPERS IN THIS ISSUE OF THE JOURNAL HIGHLIGHT SOME OF THE **HIGHLY SIGNIFICANT ADVANCES** THAT CONTINUE TO BE MADE IN COMPRESSION AND DELIVERY OF GREAT CONSUMER EXPERIENCES.

interactivity needs of the metaverse. Providing a welcome boost to future research, Mingorance also provides a software tool and methodology that other researchers can use to explore and develop power-friendly video compression techniques. It will be exciting to see the progress that can be made for existing and future video codecs as energy optimization becomes a critical performance metric alongside bandwidth and storage efficiency.

Another theme of this issue is enhancing the functionality and performance of existing video coding and distribution standards.

Markus Weber describes a way to preserve color fidelity during compression, particularly for content that contains graphics. Weber presents an RGB-based compression approach that avoids the color artifacts introduced by YCbCr compression at the same bitrate. The good news is that the approach Weber describes is based on the existing SMPTE ST 2019-1 standard.

Film grain is a hallmark of cinematic content. So much so that a film grain effect is often added in post-production to create a cinematic feel even when the content was captured digitally. The biggest challenge is that film grain is a statistically random signal that needs much bandwidth. Modern video codecs get around that by enabling film grain to be removed in the encoder and reapplied after decoding. Yet, that brings up another challenge. Most consumer smartphones, TVs, and other playback devices do not support film grain synthesis as a post-decoding process. Grois et al., describe a way to change that by enabling MPEG video codecs such as HEVC to use the film grain synthesis process required to be present on AV1-capable receivers. Their tests with professional video encoders indicate their method can result in more than 80% bitrate savings for content with heavy film grain and about 10%, which is still significant, for content with light grain.

Technologies to deliver compressed audio and video are an intimate part of end-to-end media compression. The challenge is that media is not the same as other kinds of data. That is particularly true for media over IP networks. Kostiukevych et al., identify testing of media IP networks under load as a critical step before operational full-scale deployment. Historically, that has been easier said than done, but the job should be easier using the approaches described by Kostiukevych et al. Their paper provides practical methods to sample and analyze network traffic non-intrusively using low-overhead open-source tools.

Getting ready for the future rounds out the compression theme of this issue. The following papers provide practical guidance that we could start using today to plan for new, even more spectacular products and services.

Practical guidance on using the newest international standard for video compression, Versatile Video Coding, or VVC for short, is the subject of the paper from Litwic et al. VVC is about twice as efficient as HEVC and AV1, so there is growing interest in understanding how VVC can be deployed in broadcast and stream-

1545-0279/24©2024SMPTE

ing. Litwic et al., summarize the status of VVC adoption in application standards organizations such as DVB and 3GPP and provides a glimpse into newly published guidelines from the Media Coding Industry Forum on how to use and configure some of the advancements in compression technology that are now available in VVC.

In addition to the papers on compression, two additional insightful papers are presented in this issue.

Eric Rigney guides us on creating clean soundtracks in on-set virtual production that carry efficiently to the final sound mix with minimal post-production. The challenge is that most virtual production facilities focus on creating in-camera visual effects rather than sound production. Rigney identifies improvements to sound production as more economical, efficient, and higher quality than fixing sound in postprocessing using automated dialog replacement. His paper provides a suite of practical tools and practices to create acoustic-friendly virtual production.

Footen et al., introduce the idea of a curator agent to enhance consumer experiences. A curator agent would be owned by the viewer. Its job would be a trusted personal intermediary to bring together entertainment, news, and advertisement to best match a viewer's interests. Footen, et al., identify the technologies and ecosystem evolutions we should target to enable such efficiently personalized consumer viewing choice.

The papers in this issue of the Journal highlight some of the highly significant advances that continue to be made in compression and delivery of great consumer experiences. From extending the functionality and usefulness of existing compression standards to making energy efficiency a critical performance metric to preparing to use AI and machine learning as the new normal, we can continue to look to compression to provide a versatile, efficient, and reliable foundation for our industry.

About the Author

Sean McCarthy, PhD, is currently director of video strategy and standards at Dolby, where he explores innovations, new use cases, and core technology standards that assist Dolby transform storytelling, and produce new experiences that unleash the potential of entertainment and communications. McCarthy brings a unique convergence of expertise in signal processing and the neurobiology of human vision to digital video and entertainment technology.

DOI: 10.5594/JMI.2024.YDIS4131
Date of publication: 24 January 2024

1545-0279/24©2024SMPTE

TECHNICAL PAPER

KEYWORDS HEVC // METAVERSE // MOTION ESTIMATION // SUSTAINABILITY // VP9

Green Video Compression for Metaverse: Lessons Learned from VP9 and HEVC

By Natalia Molinero Mingorance

Advanced codecs such as HEVC and VP9, integral to facilitating seamless interactions, entail considerable energy consumption, notably in resource-intensive motion estimation processes during encoding and decoding. This energy expenditure predominantly occurs within data centers during video processing. The imperative balance between visual fidelity and processing efficiency underscores the pressing need for green video compression methods.

1545-0279/24©2024SMPTE

Abstract
Over the past decade, video consumption applications have surged, reaching new heights with the metaverse's emergence. This expansion burdens networks, data centers, and devices due to increased data volume and processing, leading to substantial energy consumption and high CO_2 emissions annually. Priority should be given to developing lightweight video compression algorithms to tackle this. Current standards fall short of achieving the desired efficiency. This study conducts a comprehensive analysis of Motion Estimation (ME) in leading metaverse video compression algorithms, VP9 and HEVC. Using Matlab, an exhaustive evaluation focuses on ME, allowing an objective comparison and integrates novel sustainability assessments. The findings highlight areas for future video compression improvements, paving the way for sustainable and optimized video storage and transmission in the metaverse.

The rise of multimedia consumption, particularly video content, has become an integral part of our lives. Videos are widely used in diverse fields, such as online entertainment, education, and remote work. The metaverse, an immersive digital universe spanning virtual worlds and augmented reality experiences, is witnessing exponential growth and demands high-quality digital video to meet the increasing demand for engaging content. In spite of that, many content consumers are unaware of the growing carbon footprint associated with this paradigm shift, where video creation and communication processes contribute significantly to global energy consumption and greenhouse gas (GHG) emissions.

Within the metaverse, video content is characterized by high resolutions, real-time streaming, and interactive features. To efficiently handle large volumes of video data, two of the most commonly used video compression methods are High-Efficiency Video Coding (HEVC) and VP9.

HEVC and VP9 are cutting-edge video codecs known for their ability to significantly reduce video file sizes without compromising visual quality. These codecs enable smoother streaming experiences and faster video transmission, facilitating seamless interactions within the metaverse. However, it is essential to note that the encoding and decoding processes of HEVC and VP9, especially the resource-intensive motion estimation (ME) processes, require substantial computational power and time, leading to increased energy consumption and pollution.

Efforts are now focused on creating more energy-efficient video compression methods to address the environmental impact of video processing in the expanding metaverse. By striking a balance between maintaining video quality and reducing processing times, sustainable and cost-effective solutions can be achieved.

To address this challenge, this paper aims to help in the development of new efficient video compression algorithms prioritizing energy efficiency. Using Matlab, the research provides a standardized and freely accessible implementation for comparing the two video codecs, focusing on the ME component of HEVC and VP9. The incorporation of energy and computational complexity metrics empowers researchers to experiment with novel parameters to advance energy-conscious video compression techniques.

The contributions of this paper extend beyond codec implementation and performance evaluation. It sheds light on the importance of considering energy consumption in the design of video compression methods and its impact on the digital carbon footprint. As the metaverse's popularity grows, the development of energy-efficient video compression algorithms becomes even more critical to minimize environmental impact and reduce carbon emissions.

The structure of the paper is as follows: the next section provides a literature review of the work accomplished to date, highlighting advancements in this field and elucidating how this study can address existing gaps. The Methodology section explains the implementation of the ME processes on HEVC and VP9, provides a description of the metaverse video file used in these tests, and introduces the metrics used to assess performance in terms of energy consumption and computational complexity. The Results section presents the outcomes of predicting the same frame in both codecs, offering valuable insights into their comparative performance. In the Discussion section, a broader perspective is provided on the escalating environmental impact resulting from our collective use of digital

1545-0279/24©2024SMPTE

video, especially with the rise of the metaverse. Finally, the paper concludes by summarizing key findings and outlining promising directions for future research.

By addressing the energy efficiency challenges in video compression and providing a systematized implementation, this paper significantly contributes to the advancement of sustainable video transmission in the metaverse, bringing us closer to a greener and more optimized digital future.

Literature Review

In a study by M. Uitto,[1] the energy and power consumption of open-source video encoders, including x264 for H.264/AVC, x265 for H.265/HEVC, and VP9, were examined. H.264/AVC, where AVC stands for Advanced Video Coding, is a widely used video compression standard and the predecessor of H.265, also known as HEVC. The x264 encoder had the lowest energy consumption but the lowest compression efficiency, while the x265 encoder had the best efficiency but higher energy consumption. VP9 demonstrated a favorable tradeoff between compression efficiency and energy consumption. It is important to note that these findings are specific to the analyzed encoder implementations and may not be universally applicable to all compression algorithms. The open-source encoders were developed with different programming styles and optimization goals, making generalizations challenging. To bridge this absence, this paper offers a detailed analysis of the complexity of VP9 and HEVC, using an implementation in Matlab that provides valuable insights into the intricacies of the algorithms.

In their work, D. Grois et al.[2] conducted a complexity analysis using the reference software implementations for H.264, HEVC, and VP9. Similar low-delay configurations were employed for all encoders. The results showed that HEVC achieved average bit rate savings of 32.5% and 40.8% compared to VP9 and H.264, respectively, for 1-pass encoding. For 2-pass encoding, HEVC yielded average bitrate savings of 32.6% and 42.2% relative to VP9 and H.264, respectively. However, the VP9 encoder had significantly higher encoding times than the x264 encoder, approximately 2,000 times higher for 1-pass encoding and 400 times higher for 2-pass encoding. Notably, the evaluation was specific to the encoder implementations used, particularly without direct comparability to the previous reference. The present investigation aims to provide reliable conclusions for new algorithm designs by evaluating HEVC and VP9 through a custom Matlab implementation, allowing for the analysis of the current method's complexity bottlenecks, concentrating on ME.

R. Monnier et al.[3] presented power consumption comparisons of different available encoders (x264 for H.264/AVC, VPxenc for VP9 and its previous version, VP8, x265 for H.265/HEVC, and KVazaar for high-performance HEVC encoding), considering the Peak Signal-to-Noise Ratio for the Y luminance component (PSNR-Y) to measure the quality by comparing the original video's maximum power to the distortion or noise affecting it. They evaluated HEVC with two encoders (x265 and Kvazaar) and found that Kvazaar exhibited twice the power consumption compared to x265 when assessing PSNR-Y. While this provides insights into encoder performance, it emphasizes the importance of specific codec implementations, which may differ in programming principles, programming languages, and optimization parameters. Therefore, to accurately evaluate different algorithms, they should be developed using the same programming language and follow consistent principles, such as variable and function structures, as demonstrated in the present paper.

A. Katsenou et al. investigated the energy, quality, and bitrate tradeoffs across the following codecs:[4] Scalable Video Technology for Alliance for Open Media Video 1 (SVT-AV1), VP9 (vpxenc), VVenC (a high-performance, open-source video encoder developed by Netflix, optimized for encoding video content in the HEVC format), and x265. They proposed a new metric for the required bits: the energy cost. Similar to the previous references, their performance results were obtained using third-party implementations, introducing uncertainties. They concluded that x265 appeared to be the best choice for low-energy solutions, albeit with slightly lower average video quality. While this study helps understand the implications of input parameters on different codecs, it does not provide specific reasons for these observations or potential solutions. Moreover, the lack of detail can lead to discrepancies with the conclusions presented,[1] where VP9 is regarded as the most efficient codec. The present work isolates the ME process, the most resource-consuming part of current video compression algorithms. This way, researchers can better understand its inefficiencies and take more concise actions.

To date, there are no energy-efficient video compression publications related to video frames with metaverse characteristics.

Methodology

In this section, the implications of utilizing the same software platform, specifically Matlab, for comparing both ME algorithms (i.e., VP9 and HEVC) will be reviewed. Then, the ME principles will be described, along with the features of the metaverse video sample used for the tests. This section ends with a description of the new metrics introduced.

Using Matlab to Study Video Compression Algorithms

Comparing existing codecs using the available software is challenging and does not yield objective results, potentially leading to inconsistencies.[5] Using Matlab offers several advantages for researchers working with new video codecs. Matlab provides a clear and straightforward comparison platform, allowing for precise control of configuration parameters and features.

As images and videos can be treated as matrices of data in Matlab, inserting, testing, and analyzing new ideas in compression algorithms involving transformations, prediction, and reconstruction of matrix data becomes more intuitive.

1545-0279/24©2024SMPTE

Due to Matlab's resource consumption and execution speed when compared to compiled software in languages such as C++, the computational cost must be carefully considered during performance testing. Matlab also offers a performance testing toolkit for translating code into C and C++.

By implementing key features of video codecs, VP9 and HEVC, in Matlab, this paper provides researchers with an educational tool to easily run and understand the ME module, a crucial and time-consuming task in current video encoders with a significant impact on energy efficiency.

This unified approach allows for an objective and efficient comparison of the two encoders, enabling researchers to gain valuable insights into video compression algorithms and their implications for energy consumption.

Motion Estimation Based on Block Match Algorithms

Although VP9 and HEVC algorithms differ, both codecs employ the Block Matching Algorithm (BMA) during ME at the encoder to discover motion vectors (MVs). Moreover, ME in both codecs supports variable block sizes (i.e., 4 x 4, 8 x 8, 16 x 16, 32 x 32, and 64 x 64 pixels). **Figure 1** illustrates the BMA, demonstrating a current block from a frame compared to blocks within a specified search area in a reference frame. The algorithm determines the best match by assessing the similarity between the current block and those in the search area, identifying the MV representing the necessary displacement to align the blocks. This MV denotes the optimal position in the reference frame, facilitating motion estimation for effective video compression. The specific block of pixels selected from the reference frame, aligning with the corresponding block in the current frame as the optimal match, is termed the prediction block (PB).

As will be examined in the Results section, the resolution of MVs plays a significant role in efficiency but introduces additional complexity. Finer MV resolutions lead to increased complexity in subpixel interpolations. Additionally, representing higher-resolution MVs requires more bits. Thus, there is a tradeoff in MV resolution to balance ME accuracy for improved coding gains, the bit budget allocated for signaling MVs, and encoding complexity.

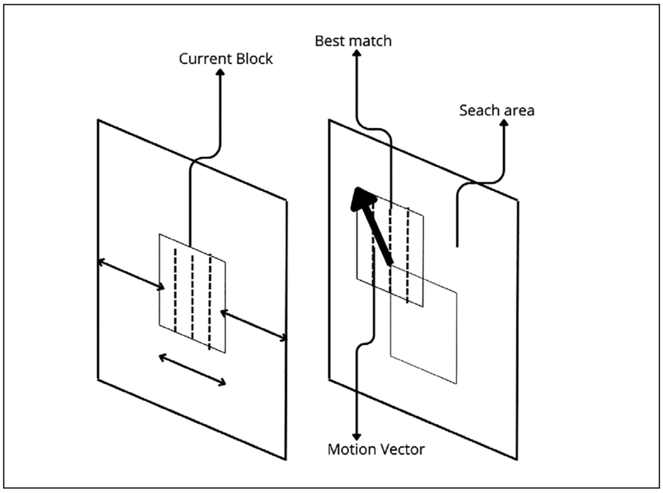


FIGURE 1. Block Matching Algorithm.

MATLAB PROVIDES A **CLEAR AND STRAIGHTFORWARD COMPARISON PLATFORM,** ALLOWING FOR PRECISE CONTROL OF CONFIGURATION PARAMETERS AND FEATURES.

Moreover, the length of the chosen interpolation filter impacts the amount of data fetched from memory and the number of operations performed. The number of multiplications and additions per sample required for interpolating the block is very significant when compared to other steps in the ME process, especially for smaller blocks of size NxM when MVs represent fractional displacements in both horizontal and vertical directions.

Both encoders employed used a 256 x 256 pixel reference and current coding unit (CU) sizes and a 32 x 32 PB size to aid testing. This PB size strikes an intermediate performance balance, as reducing it is known to increase the bit rate and decoding time, but can compromise quality.[6]

HEVC: ME Matlab Implementation

In HEVC, two search methods for ME aim to find the best-matched predicted block: Full Search and Fast Motion Search.

The Full Search method checks all points within the search window, which can be blocks of pixels or subpixels, depending on the resolution. While simple, it is time-consuming. On the other hand, Fast Motion Search checks a subset of points in multiple iterations. This method is faster than Full Search but sacrifices accuracy, making it a more common choice for software implementations.

The Test Zone (TZ) Search scheme (a Fast Motion Search method), implemented in the Matlab simulator, follows these steps:

- Square Search: it calculates the best distance (*bestDistance*) as the minimum cost among all the blocks scanned.
- If *bestDistance* > *iRaster* (typically set as 4), do Raster Search.
- If 0 < *bestDistance* < *iRaster*, do a Raster Refinement Search.

1545-0279/24©2024SMPTE

During this refinement process, subpixel interpolation of the last selected block occurs, necessitating comparisons of all pixels and subpixels from the reference frame with the current frame. This additional level of precision in the ME comes at the cost of increased computational complexity and higher energy consumption.

The complexity of an encoder is also influenced by the choice of metric used to express the similarity between a current and a reference image during ME. The most commonly used metric in HEVC is Sum of Absolute Differences (SAD). For subpixel accuracy (½- and ¼-pel), Sum of Absolute Transform Differences (SATD) is employed. SATD is more complex as it involves computing the transform of a block. Equations 1 and 2 define SAD and SATD, respectively:

$$SAD(x,y) = \sum_{i=0}^{M-1}\sum_{j=0}^{N-1}|C(i,j) - R(x+i,y+j)| \qquad (1)$$

$$SATD(x,y) = \frac{1}{2}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}|H \cdot (C(i,j) - R(x+i,y+j)) \cdot H^{T}| \quad (2)$$

In both equations, C(i, j) and R(x + i, y + j) represent pixel intensities of two images, and M and N are the dimensions of a block within the images. The coordinates (x, y) indicate the MV coordinates of the reference block, and H denotes the Hadamard transform.[7]

The Hadamard transform is a type of linear transformation with elements of +1 and -1. In the context of SATD, it is applied as a means of decorrelating the pixel values or capturing the spatial frequency information in an image or block of pixels. H is used to transform the pixel differences between two corresponding blocks before calculating the sum of their absolute values. This transformation helps capture the spatial frequency characteristics of the differences.

SAD is a simpler metric that calculates the sum of absolute differences between corresponding pixel values in two image blocks.

Both functions are fundamental operations in the block-matching ME subsystem.

VP9: ME Matlab Implementation

The implementation of the VP9 inter-prediction process in Matlab is an adaptation of the pseudo-code detailed in Ref. 8.

- Motion vector selection: finding the MV for the current block.
- Motion vector clamping: changing the MV into the appropriate precision and clamping MVs that go too far off the edge of the frame, i.e., beyond the boundaries of the frame or outside the permissible search range defined by the encoder settings. Choosing the motion vector clamping limit significantly impacts encoding efficiency, balancing accurate representation and resource allocation. The correct limit optimizes the process by excluding irrelevant data without compromising video quality.
- Motion vector scaling: computing the sampling locations in the reference frame based on the MV. The sampling locations are also adjusted to compensate for any difference in the size of the reference frame compared to the current frame.
- Block inter-prediction: obtaining the 2D array containing inter-predicted samples. The sub-sample interpolation is obtained using two one-dimensional convolutions. First, a horizontal filter is used to build-up a temporary array, and then, this array is vertically filtered to obtain the final prediction.

Among these, motion vector scaling and block inter-prediction stand out as the most demanding operations. Motion vector scaling involves computing sampling locations in the reference frame based on MVs and adjusting them to account for any discrepancies in frame size between the current and reference frames. This process requires numerous arithmetic operations and addressing calculations, significantly contributing to computational overhead. On the other hand, block inter-prediction involves sub-sample interpolation using two one-dimensional convolutions to generate the 2D array of inter-predicted samples. These convolution operations are computationally intensive, particularly for larger block sizes and high-resolution videos.

Key Features of Metaverse Videos

A typical metaverse video is a dynamic and immersive medium that plays a crucial role in shaping the user experience within the virtual environment. Several key features define the quality and realism of metaverse videos:

- Resolution: it refers to the number of pixels used to display the video image. Higher- resolution videos provide sharper and more detailed visuals, enhancing the overall immersive experience. Common resolutions include HD (720p), Full HD (1080p), 2K (1440p), and 4K Ultra HD (2160p). As the metaverse aims for realism, higher resolutions are often preferred to create lifelike and detailed virtual worlds.
- Frames Per Second (FPS): it denotes the number of individual frames displayed per second in the video. Higher FPS values result in smoother motion and reduced motion blur, which is crucial for interactive experiences within the metaverse. Standard frame per sec values are 24, 30, and 60, but for an optimal virtual reality (VR) experience, 60 frames/s or higher is recommended to ensure fluid and comfortable interactions.
- Dynamic Lighting and Shading: Metaverse videos incorporate dynamic lighting and shading techniques to simulate realistic lighting conditions. Real-time rendering of shadows, reflections, and global illumination improves the visual fidelity and adds depth to virtual scenes.
- Interactive Elements: Metaverse videos often feature interactive elements, allowing users to participate and influence the virtual environment. This could include real-time interactions with objects, characters, and other users, enabling a sense of agency and immersion within the virtual world.
- Compression and Streaming: Efficient video compression is essential for streaming metaverse content smoothly over the internet. High-quality video codecs, such as HEVC or VP9, are commonly used to reduce file sizes without significant loss in visual fidelity.

1545-0279/24©2024SMPTE

- Stereoscopic Rendering: This technique can be employed in metaverse videos designed for VR experiences. It creates a sense of depth and three-dimensionality, providing an immersive and realistic visual perception when viewed through VR headsets.

In this study, a 4K (UHD-1) raw video with a frame rate of 60 frames/sec and a duration of 120 sec served as the source material for studying ME in VP9 and HEVC compression. For this research, a single frame from the video (i.e., the same for both codecs) was selected to perform the ME analysis. The chosen frame provided a representative snapshot of the dynamic and interactive content present in metaverse environments. **Figure 2** depicts the metaverse frame utilized for these tests.

By focusing on a single frame, the study aimed to isolate and evaluate the performance of the ME algorithms employed by VP9 and HEVC in capturing the motion and temporal redundancies within the metaverse video. This approach allowed for an accurate comparison and a thorough investigation of the efficiency and accuracy of the ME process.

The frame's UHD-1 resolution ensured high detail and smooth motion representation, essential for accurately assessing the ME algorithms' capabilities. Through this targeted analysis, the study sought to shed light on the strengths and weaknesses of VP9 and HEVC's ME techniques.

Proposed Metrics Included in Motion Estimation

A combination of three metrics was used to evaluate each codec.

When evaluating video quality, metrics such as the Structural Similarity Index Measure (SSIM) play a crucial role. SSIM is a perception-based index that captures changes in structural information within an image or between different images. It takes into account important perceptual phenomena like *luminance masking* and *contrast masking*. *Luminance masking* refers to the phenomenon where image distortions are less visible in bright areas, while *contrast masking* refers to the phenomenon where distortions are less noticeable in regions with significant activity or texture. Equation 3 shows the calculation of SSIM:

FIGURE 2. Selected metaverse frame for this study.

$$\mathrm{SSIM}(\mathrm{x},\mathrm{y}) = [I(x,y)\cdot c(x,y)\cdot s(x,y)]^{\alpha} \qquad (3)$$

where x and y are the two input images compared, *l(x,y)* represents the luminance comparison between the images and captures the differences in brightness, *c(x,y)* represents the contrast comparison and takes into account differences in contrast, *s(x, y)* represents the structure comparison, capturing differences in structural information, and α is a parameter that controls the influence of each component and is typically set to 1. Each component is calculated as the average of local measurements obtained by dividing the images into smaller windows (i.e., blocks). Luminance comparison is based on mean intensity values; contrast comparison considers standard deviations and structure comparison evaluates the covariance of intensity values. Combining these components and raising the result to the power of α provides an SSIM value that measures similarity between the images, accounting for differences in luminance, contrast, and structure. Higher SSIM values suggest a higher level of structural similarity between the two images. In the test, SSIM is utilized to assess the reference PB and the current PB to determine whether ME should be computed, conserving energy when the frames exhibit significant similarities. Therefore, ME is performed when the SSIM is less than the selected threshold of 0.90.

The second metric incorporated into the VP9 and HEVC Matlab implementations is the number of computations performed in each step of the algorithms. Quantifying the complexity of the software through counting the executed operations gains a comprehensive understanding of the computational demands. It is essential to carefully assess the computational complexity of mathematical operations, particularly within programming loops, and strive to minimize resource-intensive calculations such as multiplications. Reducing this metric directly contributes to energy savings and subsequently reduces GHG emissions. Matlab's Profiler[9] proves to be a valuable tool in determining the complexity of different parts of the code, aiding in optimizing the software's efficiency.

The analysis follows by deriving operational emissions (O),[10] with a particular emphasis on the laptop's power consumption while executing each ME algorithm in Matlab. The computer utilized to do the tests was an ACER Aspire F5-573G, with the characteristics shown in **Table 1**.

TABLE 1. Laptop specifications.

Parameter	Value
RAM	16 GB
CPU	Intel(R) Core(TM) i7-7500U
CPU Frequency	2.7 GHz
OS	Windows 10
Number of cores	2

1545-0279/24©2024SMPTE

A power meter was utilized to measure the power consumption accurately. Consequently, to assess the energy consumed by the software for a specific task, Equation 4 is used:

$$O = E \cdot I \quad (4)$$

where E represents the energy consumed by the software for the task, measured in kWh. In this study, the task was defined as the execution of inter-prediction for one frame. E was determined by measuring the laptop's power consumption with a power meter, during idle state, and while running each algorithm in Matlab. The difference between these two values was then converted to kWh using the total processing time for each algorithm obtained from Matlab. The parameter I, denoting location-based marginal carbon intensity, was obtained and defined as 275 gCO_2Eq/kWh.[11] It embodies the carbon emissions associated with generating an extra unit of electricity at a specific location on the grid. This measure exemplifies the environmental impact of electricity generation in the selected region, Europe, to enhance result precision, as the tests were performed in Spain.

Results

Before conducting ME, the SSIM is obtained, yielding a value of 0.86. Since it falls below the chosen threshold (i.e., 0.90), the ME computation was carried out for the two codecs.

For HEVC-ME, detailed performance measurements were obtained for each step, including the number of computations performed and the corresponding processing time. The results are summarized in **Table 2**.

TABLE 2. HEVC-ME execution performance in Matlab.

Step	Computations	Processing Time (seconds)
Initial Search	2,753	0.068
Raster Search	548	0.029
Refinement Search	2,644	0.062

The analysis of the HEVC-ME computations yielded the following results:

- The SAD function was called 2,305 times during the Initial Search, accounting for 17% of the computing time of this search algorithm. It was called 484 times in the Raster Search, accounting for 14.3% of this step's processing time. In the Refinement Search, SAD was called 2,004 times, taking up 4.4% of the execution time for this part. SATD and subpixel interpolation were each called 64 times during the Refinement Search, constituting 33.6% and 32.9% of the total function time, respectively. Therefore, overall, SAD, SATD, and subpixel interpolation operations were the most demanding steps.
- The average power consumption was 8.1 W without the refinement step and increases to 14 W when the refinement search is included.
- The estimated operational emissions were 0.0006 gCO_2Eq without the refinement step and rise to 0.0009 gCO_2Eq when the refinement search is executed.

Furthermore, the computations and processing time for each of the VP9-ME steps were measured in Matlab, and the results are presented in **Table 3**.

TABLE 3. VP9-ME execution performance in Matlab.

Step	Computations	Processing Time (seconds)
Motion vector selection	28	0.0026
Motion vector clamping	1	0.0043
Motion vector scaling	1	0.0064
Block inter-prediction	18,471	0.072

In this case, the following results were obtained:

- The most time-consuming step is the "block inter-prediction" function, which accounted for 84.4% of the total execution time. Within this function, 47.2% of the time was dedicated to the sub-sample interpolation process.
- The average power consumption observed during the measurements was 5.1 W.
- The estimated operational emissions associated with the process were found to be 0.0004 gCO_2Eq.

The performance analysis of HEVC-ME and VP9-ME algorithms provides valuable insights into their computational requirements and energy consumption.

In HEVC-ME, the most time-consuming steps are the SAD, SATD, and subpixel interpolation functions. Particularly, the SATD function's substantial time consumption is due to the linearithmic time complexity *(nlogn)* of the absolute value of the Walsh-Hadamard transform, where *n* represents the size of the subpixel interpolated reference and current CUs. This significantly contributes to the overall execution time, resulting in higher power consumption and operational emissions.

Conversely, VP9-ME dedicates a considerable amount of time to sub-sample interpolation. This is due to VP9's utilization of an 8-tap fractional pixel interpolation filter, which enhances accuracy, although more computationally intensive than the 6-tap interpolation filter used in HEVC. Overall, VP9-ME demonstrates superior energy efficiency, leading to lower power consumption and estimated operational emissions compared to HEVC-ME.

These findings underscore the importance of understanding computational complexities and energy metrics when evaluating and optimizing video compression algorithms for improved efficiency and reduced environmental impact.

Discussion

In this study, we compared the environmental impact of video compression algorithms, specifically focusing on VP9 and HEVC. Our results revealed that VP9 emitted lower carbon emissions and consumed less power during execution than HEVC. This suggests that VP9 performs faster predicting sample frames, while HEVC operations involve higher complexity and more iterations in the search process.

To evaluate the carbon footprint of metaverse videos, we analyzed a 120-sec UHD-1 video with a frame rate of 60 frames/s. In the best-case scenario (i.e., VP9-ME), compressing each frame resulted in 0.0004 gCO_2Eq/frame emission. If this is extrapolated to the entire 120-sec video with 7,200 frames, the total estimated carbon emissions would be 2.88 gCO_2Eq. With a data center using 50% renewable energy, overall carbon emissions could be reduced to 1.44 gCO_2Eq.

Extending these findings to the vast number of metaverse users worldwide, there are approximately 400 million active users each month,[12] and assuming each user generates or interacts with an average of ten videos monthly, we can anticipate an annual carbon emission of approximately 69,120 metric tons of CO_2Eq. This carbon output is comparable to the annual emissions of hundreds of small to medium-sized power plants or tens of thousands of cars.[13] This highlights the significant environmental impact of metaverse videos on a global scale, comparable to specific industrial sectors in terms of carbon emissions. The magnitude of metaverse video emissions emphasizes the urgency for adopting sustainable practices in the rapidly expanding digital landscape.

Notably, the estimated 69,120 metric tons of CO_2Eq closely align in magnitude with the results published by Meta,[14] which state that 57,000 metric tons of CO_2Eq primarily originate from data centers processing videos and metaverse applications due to the nature of the company. This reinforces the need for increased attention to energy-efficient practices in video processing to mitigate the environmental impact of the metaverse.

It is essential to acknowledge the intricate nature of ME steps in HEVC and VP9, which leads to a non-linear relationship between the number of iterations and processing time. The complexity of the operations within each iteration is the primary factor responsible for the increased consumption of time and energy, as both processing time and energy usage exhibit a proportional connection. This emphasizes the importance of optimizing these algorithms to achieve a balance between computational efficiency and energy consumption in practical video coding applications.

Regarding the comparison of VP9 and HEVC in Matlab, although the processing time may be longer compared to other software platforms, the number of computations remains the same as it solely depends on the algorithm implementation. Thus, Matlab allows for a fair comparison of both codecs.

Given the high volume of videos today, even if the pollution were cut in half by using faster software, the associated emissions would still be a concern. This highlights the need for more energy-efficient video compression algorithms that use real power consumption data and metrics like Operational Emissions (O). These metrics can be employed during algorithm development stages, where real measurements can be taken and specific hardware data can be obtained

Join the Board of Editors
Volunteer to help shape and maintain the Journal's high editorial quality.
MI
MOTION IMAGING JOURNAL
SMPTE

from suppliers for accurate calculations. Furthermore, these metrics can be incorporated into the algorithms as decision values to optimize efficiency.

In line with the current investigation, perception-based metrics such as SSIM can be employed to assess the resulting compressed video and be integrated into the compression algorithm. This integration optimizes the balance between perceived quality and computational considerations, including energy consumption. In regions characterized by high SSIM values, indicating substantial structural similarity, it becomes possible to skip the ME process, thereby reducing algorithmic complexity. Furthermore, in scenarios where CUs exhibit significant similarities, strategies like aggressive quantization, which reduces bit rate allocation without compromising perceived quality, could come into play.

Analyzing the efficiency of current algorithms, as demonstrated in the Matlab code developed for this work, is an essential initial step towards a more sustainable metaverse.

Conclusion

The rapid growth of mobile devices, internet accessibility, video-on-demand services, social media, and the emergence of the metaverse, have led to a significant surge in digital video traffic. However, this growth has also increased processing complexity in current compression methods, resulting in higher energy demand and associated pollution.

Addressing this issue requires lighter compression methods that can efficiently handle video encoding. This study compared the performance of VP9 and HEVC, focusing on their ME processes, using a custom implementation in Matlab for a metaverse video sample frame. The findings revealed that both codecs employ thousands of computations to predict a single video frame. While both codecs possess remarkable features and designs, the results underscore the urgency of developing more efficient and sustainable video compression techniques. Reduced video processing times and smaller file sizes enable quicker upload, download, and streaming, enhancing user experience and reducing overall energy demand on servers, data centers, and network infrastructure.

Future research of this work will address inefficiencies in the full compression algorithms, optimize the entire compression pipeline, and promote eco-conscious behavior among users. The overarching objective is to establish criteria for novel video compression algorithms that combine ecological sustainability with superior experiential quality. To facilitate this pursuit, the Matlab code for VP9 and HEVC is accessible through the Harmony Valley research project's website.[15]

The code for two Android apps dedicated to video compression and "eco-cam" functionality has also been shared. This initiative invites enthusiastic researchers to partake in the collective endeavor of advancing video compression standards with heightened efficiency. Energy consumption will be a primary design constraint in technological advancements, ensuring that people can enjoy digital technology without causing harm to the planet. Prioritizing energy efficiency in video processing for the metaverse and other digital applications is crucial. By optimizing video compression techniques and reducing computational complexity, we forge the way for a greener digital landscape.

Acknowledgments

The author wishes to thank the European University of Madrid and Banco Santander Universidades. This work was supported in part by a grant from the Singular Alumni Award.

References

1. M. Uitto, "Energy consumption evaluation of H.264 and HEVC video encoders in high-resolution live streaming," Proc. *2016 IEEE 12th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, pp. 1-7, New York, NY, 2016, doi: 10.1109/WiMOB.2016.7763234.
2. Dan Grois, Detlev Marpe, Tung Nguyen, Ofer Hadar, "Comparative Assessment of H.265/MPEG-HEVC, VP9, and H.264/MPEG-AVC Encoders for Low-delay Video Applications," *Proc. SPIE 9217, Applications of Digital Image Processing XXXVII*, 92170Q, 23 Sep. 2014; [Online]. Available: https://doi.org/10.1117/12.2073323
3. R. Monnier, K. Jerbi, and M. Uitto, "Specification of Power Efficient Encoder-Transcoder," Sep. 2017. Accessed Dec. 23, 2022. [Online]. Available: https://convince.wp.imtbs-tsp.eu/files/2017/09/CONVINcE-D2.2.2-Updated-specification-of-power-efficient-encoder-V1.01.pdf
4. A. Katsenou, J. Mao, and I. Mavromatis, "Energy-Rate-Quality Tradeoffs of State-of-the-Art Video Codecs," *Electr. Eng. and Syst. Sci., Image and Video Processing*, Oct. 2022.
5. T. Laude, Y. G. Adhisantoso, J. Voges, M. Munderloh, and J. Ostermann, "A Comprehensive Video Codec Comparison," *APSIPA Transactions on Signal and Information Processing*, Vol. 8, Nov. 2019, doi: 10.1017/atsip.2019.23.
6. J.-R. Ohm, G. J. Sullivan, H. Schwarz, T. K. Tan, and T. Wiegand, "Comparison of the Coding Efficiency of Video Coding Standards—Including High Efficiency Video Coding (HEVC)," *IEEE Trans. on Circ. and Syst. for Vid. Tech.* 22 (12):1669–1684, Dec. 2012, doi: 10.1109/tcsvt.2012.2221192.
7. A. M. Joshi, M. S. Ansari and C. Sahu, "VLSI Architecture of High Speed SAD for High Efficiency Video Coding (HEVC) Encoder," *Proc. 2018 IEEE International Symposium on Circuits and Systems (ISCAS)*, Florence, pp. 1-4, Italy, 2018, doi: 10.1109/ISCAS.2018.8351271.
8. J. Hunt and A. Design, "VP9 Bitstream & Decoding Process Specification," Mar. 2016. Accessed: Nov. 15, 2022. [Online]. Available: https://storage.googleapis.com/downloads.webmproject.org/docs/vp9/vp9-bitstream-specification-v0.6-20160331-draft.pdf
9. Mathworks, "Run code and measure execution time to improve performance - MATLAB - MathWorks España," mathworks.com. Accessed Jan. 05, 2023. [Online]. Available: https://mathworks.com/help/matlab/ref/profiler-app.html
10. G. S. Foundation, "Software Carbon Intensity Standard," GitHub, Nov. 01, 2021. Accessed Dec. 16, 2023. [Online]. Available: https://github.com/Green-Software-Foundation/sci
11. "Greenhouse gas emission intensity of electricity generation — European Environment Agency," Jan. 26, 2023, www.eea.europa.eu. Accessed Dec. 20, 2022. [Online]. Available: https://www.eea.europa.eu/data-and-maps/daviz/co2-emission-intensity-12#tab-googlechartid_chart_11
12. Metaverse of Things, "Metaverse Users Worldwide – Metaverse Statistics to Prepare for the Future," April. 1, 2023. Accessed June 1, 2023. [Online]. Available: https://metaverseofthing.com/metaverse/metaverse-users-worldwide/
13. U.S. Environmental Protection Agency (EPA), "Greenhouse Gas Equivalencies Calculator." www.epa.gov, Aug. 28, 2015. Accessed July 1, 2023. [Online]. Available: https://www.epa.gov/energy/greenhouse-gas-equivalencies-calculator#results
14. Meta, "2021 Sustainability report." Accessed Jan. 15, 2023. [Online]. Available: https://sustainability.fb.com/wp-content/uploads/2022/06/Meta-2021-Sustainability-Report.pdf
15. N. Molinero, "Harmony Valley – Research Project Resources, [Online]. Available: https://linktr.ee/harmonyvalley

About the Author

Natalia Molinero Mingorance, a telecommunications engineer, is balancing her research and development commitments in Harmony Valley with her new role at Airbus Defense and Space, where she has the opportunity to explore her other interests, such as wireless communications and cybersecurity.

Presented at the 2023 Media Technology Summit. Hollywood, CA, 16-19 October 2023.

DOI: 10.5594/JMI.2024/EBVF7405
Date of publication: 24 January 2024

1545-0279/24©2024SMPTE

KEYWORDS VC-3 // RDD 50 // RGB COMPRESSION // ALPHA CHANNEL // TUNABLE BITRATE

Avid DNx GX

A High-Quality, Flexible RGB(A) Codec at Commodity Bitrates, Combining SMPTE ST 2019-1 (VC-3) and SMPTE RDD 50 (DNxUncompressed)

By Markus Weber

Abstract

When the tunable compression feature from the QuickTime Animation codec was withdrawn, it left behind a gap in its wake, which animators have been struggling to fill with alternatives. The natural first choice, commodity video codecs, leads to color deterioration at sharp edges when applied to, e.g., broadcast graphics, virtual productions/volumes, VFX exchange, augmented graphics, or motion graphics. The root cause can be found using YC_BC_R sub-sampling as part of the compression, even if the original imagery was in RGB. These artifacts are usually acceptable for direct graphic transmission. However, not if further image processing (warping, blending, compositing) is required. The artifacts can be avoided by using an RGB-based compression approach. Largely unknown, the 444 mode (12/10 bit) of SMPTE VC-3 (ST 2019-1), which supports RGB-based compression, is adjustable to target bitrates down to about 20:1 compression. A minor modification can also be applied to 8-bit levels for even better quality at the same bitrate. Compared to YC_BC_R 4:2:2 at the same target bitrate, the compressed images show vastly superior visual quality. The compressed results can be packed into a DNxPacked bitstream format, which is defined in SMPTE RDD 50 Part 1. This allows combining the RGB compressed filler with an RLE compressed alpha channel of completely independent bit depth, bypassing the VC-3 limitation requiring the alpha channel to use the same bit depth as the filler.

Image codecs have existed since the early 1990s (JPEG was originally published in 1992). They form a subgroup of compressors that focuses on photographic images, exploiting human vision system (HVS) characteristics to achieve compression ratios far exceeding those of entropy-based compressors. Modern video codecs evolved from image codecs by additionally taking advantage of inter-frame temporal and spatial redundancies to reduce the compressed bitrate over the (display) time.

Mezzanine (production) video codecs usually forego the inter-frame, correlation-based compression methods and stick to I-frame compression schemes. There are multiple reasons: Speed of access and consistent image quality are just some of them. Intra-frame codecs therefore, share more attributes with image codecs than their highly specialized distribution video siblings.

All image codecs achieve higher compression ratios by making tradeoffs between the partial signal loss incurred in the coding process and the visibility of the associated artifacts to the viewer. The goal is to realize the best possible compromise between achieved visual quality and the given bitrate. These two goals are generally considered antagonistic and mutually contradictory.

The traditional I-frame compression approach generally follows, at a very high level, a 4-stage process:

1. Select the smallest acceptable uncompressed[i] input format as the starting point.
2. Apply a set of transformations that try to separate large-scale image correlations from small-scale image detail. The former can be stored as a summary value, while the latter needs to be stored individually.

[i] Packed formats like 4:2:2 are not considered “compressed.”

1545-0279/24©2024SMPTE

3. Introduce a *lossy* coding scheme for the transformed values, focusing loss on the small-scale image detail while retaining the larger-correlations.
4. Apply a *lossless* coding scheme for storing the residual values.

Modern (inter-frame) codecs push the limits for better coding results by primarily adding massive complexity in steps 2 and 3 (this includes using AI technology) to expose even more correlations and, to a much lesser degree, using new methods for the final stage 4.

Unless for very special images, which allow you to exploit coding schemes like RLE coding, stage 4 is generally achieved as some entropy coding scheme. However, the compression achievable in this stage is limited by the so-called Shannon limit, named after Claude Shannon, who in 1948 proved that a lossless compression scheme based solely on the probability of an arbitrary symbol alphabet can only asymptotically approach but not exceed, a hard limit, which is directly derived from this probability distribution. Improvements in this area are, therefore, necessarily small and modern methods like Arithmetic Coding or Asymmetric Numerical Systems (ANS) represent just small improvements over the previous generation's result. The significant improvements associated with these schemes are due to additional methods applied prior to or in conjunction with entropy coding, such as Context Adaptive Coding in CABAC.

All image codecs achieve higher compression ratios by making tradeoffs between the partial signal loss incurred in the coding process and the visibility of the associated artifacts to the viewer. The goal is to realize the best possible compromise between achieved visual quality and the given bitrate.

The one area that is typically taken for granted is stage.[1]

Preserving bandwidth is not an objective that arrived on the stage with video compressors. Even uncompressed video is (usually) already compressed, despite the fact that users frequently forget this:

The use of color difference signals (YC_BC_R), combined with sub-sampling, is already an image-specific compression technique, which goes back to the introduction of color TV. When transmission standards for color television were developed in the 1950s and 1960s, the signal had to be backwards compatible with the then-dominant black-and-white (B/W) television sets. This mandated that the B/W signal was retained in the color signal (Y), with the additional color information squeezed into the higher end of the signal's bandwidth spectrum, where its introduction would not cause significant (visual) disturbances to the signal processed by the B/W receivers. The color signal had to be "compressed," to accommodate the lower available bandwidth, respectively sub-sampled as color difference signals (U and V). This significantly reduced the size of the transmitted signal compared to the RGB representation, which is derived directly from the color primaries.[ii]

Creating a YC_BC_R signal employs one extra transformational approach (RGB → YC_BC_R) to expose a strong signal correlation (the luminance part isolated in Y). This is then combined with a stage 3 lossy technique (sub-sampling), the characteristics of which are based on the specifics of the HVS as follows:

The HVS is known to resolve brightness details much better than color details. Sub-sampled YC_BC_R signals utilize this to dedicate more bandwidth to the Y channel and less to the two chroma channels.

Image codecs then typically exploit this a second time: They start with a sub-sampled YC_BC_R signal and then dedicate more bandwidth to the Y channel than the two chroma channels during the stage 2 and 3 processing. The collective argument for lossy compression is that it is deemed okay as long as it does not become instantly recognizable to the viewers.

The Graphics Problem

Graphics, primarily broadcast graphics, form a sub-set of images that can also be considered video content when present in the form of motion graphics. They are, however, associated with several unique traits that single them out in the large "image/video content" family.

Graphics always originate on a computer. They are produced by some form of computing process (also known as a "a program") and usually with a lot of human interaction. They are viewed, manipulated, and assessed on a computer, using a computer monitor during the creation process. Unlike video monitors, computer monitors are based on (4:4:4 sampled) RGB input, and the graphic editors are used to modify colors using RGB controls.

Unless textures are mapped to part of the graphics, large parts of the graphics are frequently formed by monochromatic (or what a human viewer would consider monochromatic) swatches with precise vertical and horizontal edges. Plus, there is practically no noise present in the signal.

These specific traits can make graphics a good candidate for other compression algorithms e.g., run-length encoding (RLE), which utilizes these characteristics to establish large-scale correlations by very simple analytical

[ii] It is worth noting the fact that graphics are "real" RGB, while video signals typically are acquired via an image sensor using a Bayer pattern, which happens to consist of R, G, and B sensitive "pixels" – but only 1 channel per spatial position. To get to the same starting point as graphics (4:4:4 sampling) they RAW Bayer patterns have to be subject to elaborate mathematical processing, which will interpolate/infer the two missing color channels for each pixel.

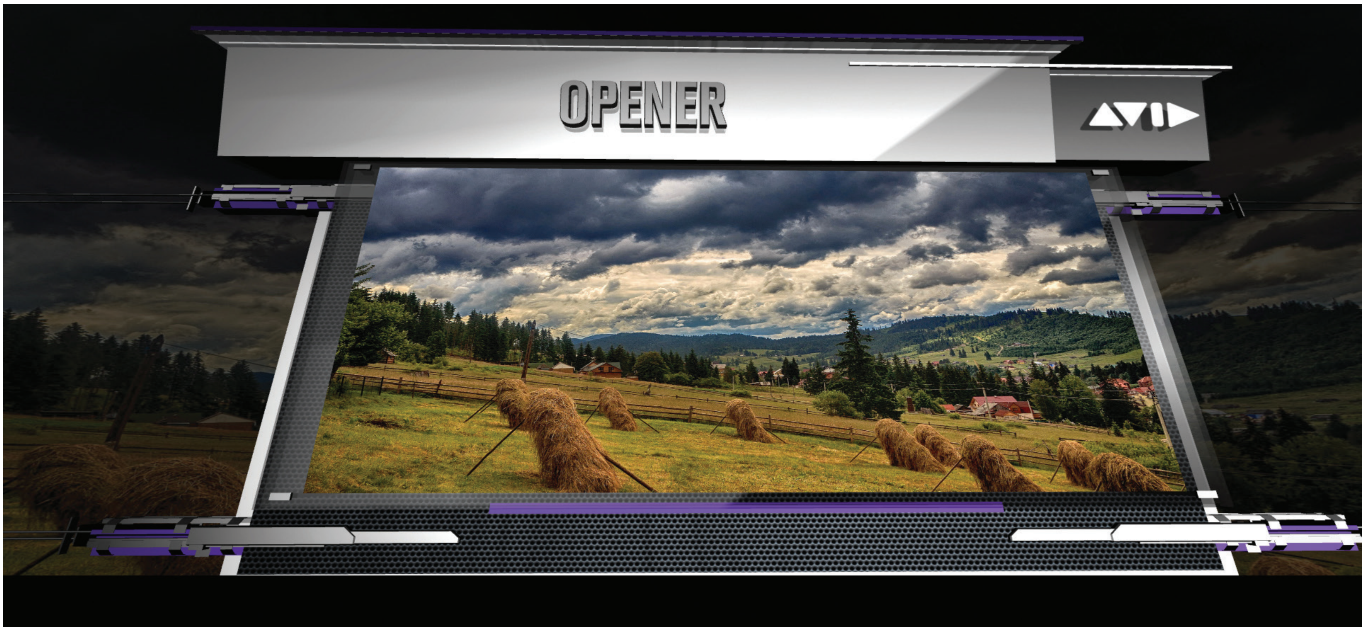


FIGURE 1. An example of a modern broadcast graphic, which still retains the geometric attributes of a graphic, but has lost the "large-scale monochromatic" attribute, associated with graphics in the past, even though a human observer would readily identify monochromatic areas in the graphic.

means, allowing them to blow through the entropy-based Shannon limit and achieve high compression ratios.[iii]

These advantages are, however, tenuous at best: The effectiveness of RLE compression, for example, stands and falls with the repetition of the same color value for long stretches of the image (the "color swatches"). Once that trait is lost, RLE compression can quickly result in images that are larger in RLE "compressed" form than their "uncompressed" originals. Also, while the monochromatic color swatches were the standard 20 years ago, today's standard is elaborately shaded 3D models, resulting in rich gradient and hue variations, which a (lay) human observer visually interpret as "monochromatic," even if it technically isn't.

This teaches an important lesson: Technical reality and human perception are two very different things:

The human brain is used to "interpret" the raw data from the eyes. One (standard) example to demonstrate this is showing an observer a white sheet of paper and asking: "Is this white?"

Depending on the accommodation time of the eye, the answer will be overwhelmingly "Yes," irrespective of illumination source. The human brain will automatically consider surroundings, interpret the raw data, and, as a result, ignore/average out variations.

If you ask a lay human observer if large parts of the "OPENER" in **Fig. 1** are monochromatic, the answer will still be "Yes." The human brain automatically establishes a "context" and thus associates the gradients and edges with artifacts of where the illumination sources are placed and how the (perceived) 3D features create shadows. As a result, it subconsciously filters out anything that is not meeting its "acquired interpretation pattern" – at least until the deviations become large and counter-intuitive.

Technically, however, the image has almost no identical colors, and the interpretation of 3-dimensionality and shadows is "pure human imagination," based on long-term acquired evolutionary knowledge.

This technical loss of monochromatic content in modern graphics has effectively led to a compression crisis: One of the previously most widely used formats for graphics was the Apple Animation codec, which supported lossless and lossy compression and an alpha channel. When Apple discontinued supporting this on their newer platforms, it forced users to look for suitable alternatives. They were forced to revert using uncompressed storage during exports, using single image formats like TIF, TGA, PNG, DPX or OpenEXR, which, for the most part, are not using any compression or are available as a video (stream) codec but are widely supported by both graphics and video editing applications (in the form of image sequences).

While the "flatness" of graphics has been replaced with rich graduation, the geometric, color-choice, and color-edge aspects of graphics frequently haven't changed dramatically.

Graphics still retain very sharp edges, which are frequently defined along hue boundaries, not luminance boundaries. Additionally, those edges often run exactly along rows or columns of the pixel grid. Colors for graphic elements are also frequently selected to provide strong color contrast in addition—or as an alternative—strong luminance contrast, favoring highly saturated, non-natural, and frequently complementary color choices that clearly stand out from any natural image content, over which they will eventually be layered.[iv]

[iii] RLE does not really overcome the Shannon limit: The RLE codes themselves form an alphabet (a different one) with a non-uniform probability distribution, which can be entropy-coded itself to achieve an even higher storage efficiency. The VLC coding of ST 2019-1 is partially based on this approach.

1545-0279/24©2024SMPTE

These attributes are not associated with images captured from real-life scenery via a camera. Natural content is usually only mildly saturated, object boundaries are typically associated with solid luminance edges, and the characteristics of the capture system ensure that edges very rarely align precisely horizontally or vertically. The physics of the lens/aperture system of the camera already introduce a slight blurring/antialiasing effect at any of the edges, and the surrounding environment will ensure, through the impact of (indirect) illumination and the material's reflective properties, that edges in natural images are almost always (color) anti-aliased. Exceedingly sharp contrast is almost always a sign of secondary image processing/enhancement, which may be part of the creative intent.

The influence of these camera/reality-induced traits of real-life imagery on the effectiveness of video compression are very often overlooked. But they become quickly apparent if one tries to solve the graphics compression crisis by employing bitrate-tunable video codecs to overcome it.

The Codec Produces Wrong Colors...

Given the large size of uncompressed graphics, it is natural that users have started to look for alternatives, preferably commodity codecs, which allow you to dramatically reduce the size of the stored content while keeping the image quality high. The obvious choices are the typical I-frame mezzanine codecs like Avid DNx (VC-3), Apple ProRes, Sony XAVC Intra (H.264), or J2K, which are widely used in production today. These codecs all have an excellent reputation in balancing between quality and compressed size. Until recently, the first three outdistanced J2K when it came to performance, making them the only reasonable choices during the editing process. They also differ in the choice of the fundamental compression technique.

While the first three can trace their roots back to JPEG and the DCT-based compression scheme it introduced, J2K takes an entirely different approach based on multi-scale Wavelet compression, which, based on its nature, could efficiently be processed only on dedicated specialized hardware or massively parallelized architectures as found on graphical processing units. The new J2K HT might change the latter part; however, for further discussion, we will limit the field to VC-3 and ProRes as typical examples, as XAVC-I is just a 1-frame-long group of pictures (GOP), which happens to consist of only an I-frame, primarily by implication, not design.

Of course, these codecs are video, meaning that most of their profiles are based on sub-sampled YC_BC_R. To use them for graphics, you must convert your RGB graphics to the YC_BC_R representation and then reduce the color resolution via the sub-sampling process.

VC-3 (SMPTE ST 2019-1)[2,3] and ProRes support a high-end 4:4:4 mode without sub-sampling. However, this mode is associated with a very high resulting bandwidth, and it is associated with an input bit depth of at least 10 bits. These modes target prime content, where quality is the first consideration and bandwidth is the second.

In the reference case, which led to the approach discussed here, the best possible quality was not the driving factor. The context was a daily live breakfast broadcast, highly branded and produced in a virtual set, with thematic inserts on a (virtual) video wall in the background. The source of the insert was either camera footage or headliner graphics, with the latter produced in After Effects and then delivered as stored footage for playback during the live event.

An interaction with the editors working with the graphics material resulted in a request to reduce the compressed bitrate by at least a factor of 5 compared to the uncompressed original.

To achieve reasonable compression, the graphic editors tried to use Avid DNx and Apple ProRes on one of the lower bitrate levels, which are associated with 4:2:2 sub-sampling. The rationale was that JPEG, as their standard and more primitive ancestor, is routinely used to store screenshots and other graphic material if you need to preserve bandwidth.[v] The results of these experiments were highly unsatisfactory.

Figure 2 shows the uncompressed original. Note that variations seen in the teal background, when looking closely, are real and caused by a partial transparency of graphic overlay, where the natural background contents (winter road scene) are showing through.

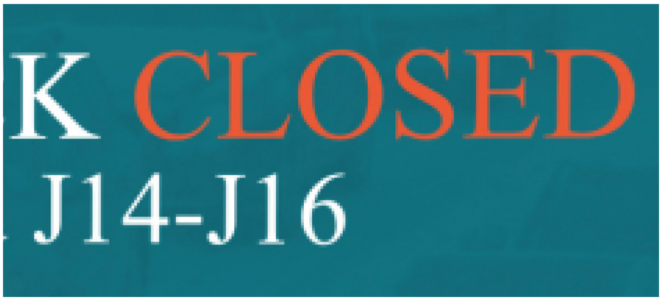


FIGURE 2. Detail of the graphic—uncompressed RGB original.

When compressed with one of the (bitrate) economic modes of ProRes or DNx (e.g. SQ), the result is shown in **Fig. 3**.

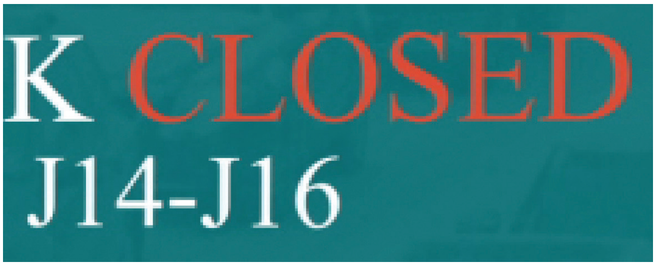


FIGURE 3. Detail of the graphic—compressed with Avid DNx HQ or ProRes 422 HQ.

[iv] The only exception is CGI (Computer Generated Imagery), which tries to reproduce natural color and luminance compositions/textures usually as seen through a camera's lens.

[v] JPEG is always YCBCR 4:2:0 sub-sampled, because it was developed for digital (still) cameras, meaning interlaced signals were never an issue.

1545-0279/24©2024SMPTE

While the white text displays more or less correctly, the orange text appears to be heavily associated with coding artifacts, resulting in what the graphic editors termed "wrong colors."

The Deeper Look

The analysis of the problem can be summarized as follows: It is a cod*ing* artifact, but it's not a co*dec* bug.

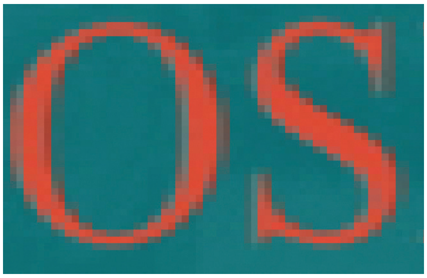


FIGURE 4. Magnification detail of the coding artifacts.

The Teal and Orange colors were (likely consciously) chosen as complementary colors by the graphic editors (**Fig. 4**). The Orange and Teal are virtual point mirrors in the C_B/C_R plane of a vectorscope. Both have a very similar luminance value Y associated with them, and both are dominantly C_R-heavy, with only a small C_B contribution **(Fig. 5)**.

The codecs differentiate their quantization settings only between the Y and the Chroma channels; not between C_B and C_R. They will allocate more resulting bandwidth to the Y than the two Chroma channels, meaning chroma usually gets compressed more than luma.

The White text survives the compression largely intact because there is a strong Y difference between the White and the Teal (the White is almost twice as bright as the Teal), with practically no chroma information coming from the White. Thus, the codecs can allocate significant bandwidth to preserving the (already favored) luminance-defined edges and must expend only very little bandwidth on the remaining Teal contributions, which are already small and require minimal bandwidth or compression. The result is a very good coding quality for the White text.

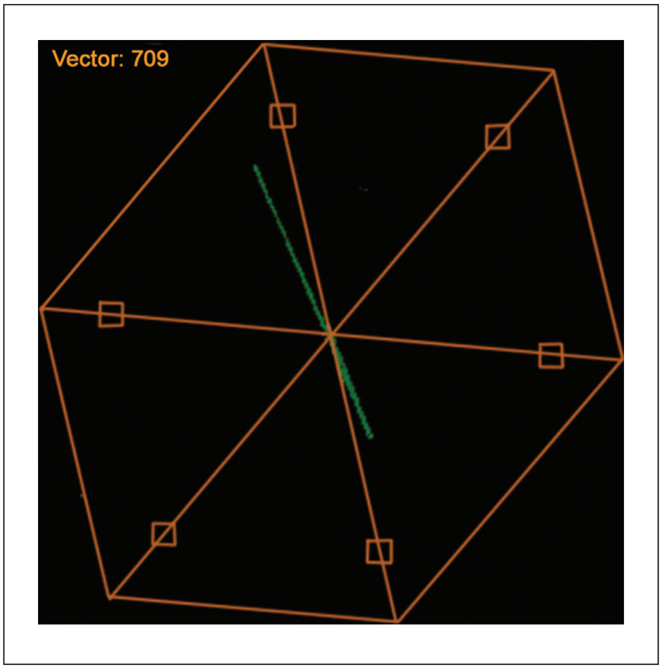


FIGURE 5. Vectorscope view of the image detail shown in Fig. 5.

For the Orange/Teal edges, however, there is very little Y difference between the neighboring colors. DCT-based codecs essentially apply a low-pass filter to the video content during compression, discarding high-frequency components to preserve bandwidth. However, the high-frequency components are what defines the "crispness" of the image. Removing them has the same effect as running a blur filter over the image.

All the DCT-based video codecs, unlike their JPEG predecessor, do not apply constant quantization pressure across the entire image. They divide the image into much smaller sections for which the strength of the quantization pressure can be adjusted separately. Of course, this extra subdivision effort is conducted in the interest of preserving the best image quality possible at the given bitrate.

The codec will preferably apply less quantization pressure in areas with lots of detail (many high-frequency components) and apply lots of compression pressure in largely uniform areas (few high-frequency components). Detecting and controlling this distribution to create a detail- rich and high-fidelity image is the art of the encoder.

In the case of the graphics problem, this strategy starts to backfire heavily: The codec will dominantly look at the Y channel when choosing the quantization during the encoding process. As the Orange/Teal edges are relatively flattish in the Y, the codec is induced to use a high compression setting because eliminating the small high-frequency components in the Y channel does not seem to alter the decoded Y results dramatically.

However, this will then spill over into the chroma channels, which will receive the same quantization pressure as the Y channel. The chroma channels will receive significantly more quantization pressure than you would normally receive for such a circumstance—simply because almost no edge is detected in the Luma.

So, what about the edge definition in the chroma?

Inspecting **Fig. 2**, the edge seems to be clearly defined and visible to a human, and by all rights, a codec modeled for the human visual system should detect and handle this properly. Still DNx and ProRes 4:2:2 profiles,[4] both fail in the same way, completely independently.

What is easily overlooked is that **Fig. 2** shows the RGB (4:4:4 sampled) original, rather than the 4:2:2 sub-sampled codec input. The general geometric attributes of graphic content reappear with a vengeance here.

When inspecting **Figs. 3** and **4**, you notice that the visible artifacts are present dominantly at the horizontal edges, but not the vertical ones. The codecs perform a proper job vertically but not horizontally. This provides the critical clue that the issue is introduced by the 4:2:2 sub-sampling.

If the Orange/Teal edge falls between 2 pixels whose

1545-0279/24©2024SMPTE

chroma values are merged during sub-sampling, the resulting color will be (more or less) a 50:50 mix of the neighboring pixel's chroma values. As the colors are effectively complementary, the resulting sub-sampled color will essentially be Grey. This is an important observation.

To the codec, this will look like two color areas separated by a clearly defined, 1 sample (2 pixels) wide grey boundary, and as the codec tries to preserve detail when it sees it, it will dominantly focus on preserving this grey boundary at the expense of the neighboring, more uniform color areas (**Fig. 6**).

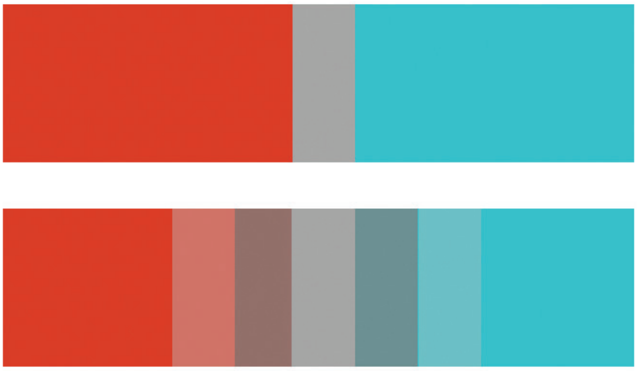


FIGURE 6. The effect of low-pass "ringing," spreading the 2 pixels Grey boundary to many neighboring pixels, making a previously almost non-noticeable artifact highly visible.

As a result, the codecs perform as intended, but are limited by the quality of the input data. Before the signal got to the codec, the 4:2:2 sub-sampling had already done the damage.

As stated in the introduction, this is an unusual occurence in real-life camera footage. The issue is avoided by either the limitations of the camera optics (blurring the color-edge definition due to the *sinc* filtering characteristics of the lens/aperture system) or the reflective properties of the material and its surroundings. Even if such a circumstance occurs in two neighboring pixels, it is extremely unlikely that the camera will be so perfectly aligned and stable that a true vertical edge spanning many pixels is obtained. In other words, even if such a case occurs in camera footage, it won't be visible/noticeable to an observer because it is either too small or the human brain will filter this out as inconsistent. But graphic content completely breaks these implied filtering mechanisms.

Mezzanine vs. Distribution

End-user distribution occurs practically exclusively in sub-sampled YC_BC_R. Once a graphic reaches this stage, the sub-sampling problem will inevitably occur. However, the audience in this case are lay viewers, and will likely not notice the problem. Plus, the problem won't get further amplified or spread, as there is no more secondary processing. The distribution encodings (Long-GOP) will also offer a better quality (less compression) in the leading I-frame image of a GOP, esp. in the case of slowly moving graphics, and thus are less affected by the emphasizing effect of high quantization pressure. But any damage that was already done during the production phase can certainly not get remedied at this stage anymore.

Much broadcast graphic content is done live, practically as the last production step. The graphics system will generate the content as uncompressed RGB and blend it directly with the live video signal. This blending may be achieved either as a simple keying operation or through a more complex, 3D-model-based, texture mapping approach, as shown in **Fig. 1**. Regardless of which approach is used, this will eliminate transformations of the graphics as the root cause for any of the issues mentioned.

However, this argument breaks down if some elements contributing to the final composite must be pre-rendered and not generated live. This was the case, for the graphics insert in the reference case mentioned above. In this case the graphic content needs to be stored in some mezzanine format, which does not compromise/constrain the usage in the final composition process. While some use cases require storing the alpha channel with the video, it can also be a video without alpha. The issue outlined in this paper is about the video generated by graphics systems, not the alpha-channel compositing.

The intention is to get the lowest bitrates in the mezzanine format the editor can pass without running an expensive final conform.

In the reference case used in this paper, the graphic was authored in After Effects and represents a compositional intermediate. The downstream component (a virtual set) picked up the graphic as a "live texture." It was subjected to additional processing steps (tilt and rotate), which may completely change the pixel positions and the relative layout. According to the laws of error propagation, errors never cancel out; the more processing they undergo, the larger the errors will become. By the time the signal gets to the final distribution stage, an already bad quality can spread so that even a lay user will start to notice the signal deterioration.

VC-3 444 Level

One of the major advantages of VC-3 is that its 444 level supports both RGB and YC_BC_R 4:4:4 input. The user can determine if the level should use RGB or YC_BC_R exclusively or if it wants to let the codec determine the chosen approach on a macroblock basis (Alternate Color Space encoding).

It may sound counterintuitive to increase the initial footprint to get a better quality at the end, and especially at the same bitrate. However, the 2016 revision of VC-3 introduced a couple of minor modifications that allow the 444 level to achieve much higher compression ratios without violating the VC-3 compliance constraints and, as it turned out, to even provide a better visual quality in almost every respect.

Part 1: Variable Bit Rate Mode

VC-3 typically uses constant bitrate (CBR) mode to achieve a constant frame size of the content and forego having to support an index in the storage files. This will speed up access to the essence during read processing, as no index needs to be read, parsed, and tracked (which may become a major factor if hour-long recordings are used, as found frequently in broadcast-based acquisition (e.g., sports).

The introduction of lossless alpha support in 2016 was accompanied by the addition of a variable bitrate (VBR) mode. Run-length encoding (RLE) compressed alpha may vary primarily in compressed size, so mandatorily retaining the CBR mode made no sense because it would counteract the RLE objective. The only provision for VBR mode is that the resulting compressed frame sizes cannot exceed the CBR frame sizes.

Part 2: Adjustable Output Frame Size

The encoder's rate controller is not dependent on the target compressed frame size. It still needs to adjust to varying input complexity. It does not matter if an increased quantization pressure originates from a complex input or a lower target bitrate. As a result, the RGB mode can operate at a substantially lower target bitrate than defined for the CBR case. The bitstream must simply be marked as VBR.

Setting the VBR flag does not imply in any way that the resulting compressed frames will wildly vary in frame size. The VBR mode in VC-3 is defined opportunistically: if the content is so simple that even zero quantization pressure cannot fill the allocated bitrate budget (for example, monochrome input), then the codec does not need to artificially inflate (pad) the content to reach the (constant) compressed frame size limit. However, it does not prevent the insertion of padding if this is desired.

In VBR mode, the codec can fully fill the specified target bitrate. If you drive the encoder with a lower target bitrate, you will still get practically constant frame sizes all the time; covering the balance via padding is an encoder implementation option—entirely within the limits of the VC-3 spec.

Part 3: Delayed Bit Depth Limiting

Figure 7 shows a block diagram of the signal processing during encoding in Avid DNx, which is one of the implementations of the VC-3 standard. Note that the standard specifies only the decoding process, not the encoder.

In VC-3, the inverse discrete cosine transform (IDCT) was intentionally defined mathematically and not as an integer implementation. Of course, in practical implementations, the discrete cosine transform (DCT/IDCT) will be implemented with some fixed-point arithmetic. But the fixed-point arithmetic and its precision aren't part of the standard. RP 2019-2 specifies an entirely optional minimal precision constraint for the integer implementation of the DCT at given bit depths. Implementers are, therefore, free to process all inputs at a higher precision than the minimal precision required without losing the "compliant" attribute to their implementations.

For a hardware implementation, this may make a lot of difference in terms of costs, but for a software implementation, which today runs on 64-bit architectures, this is (almost) a moot point.

Contrary to widespread perception, the different levels in VC-3 do not specify an input bit depth for the DCT. As has been clarified in Amendment 1 to ST 2019-1, the constraint is the design (input) bit depth at the start of the quantization stage (ST 2019-1, Section 8.2.8.3, Table 14). An implementation can, therefore, postpone handling the input sample bit depth to the point marked as "quantization range normalization" in **Fig. 7**.

This approach is normally used to allow processing of higher bit depth input (10 and 12 bit) at the 8-bit design bit depth levels without clipping the input precision up-front. Respecting the higher precision of the input will slightly alter the relative values of the DCT coefficients (non-uniform), providing a more accurate coefficient representation compared to up-front clipping. While intended for handling the "downsample" case, the approach works in both directions if implemented this way. You can also feed 8-bit RGB-based input to the 444 level. It won't give you any precision gains, but it is not prohibited either.

This ensures that all realistically encountered RGB representations can be encoded without violating the VC-3 standard. The results will just not be optimal for 8-bit input.

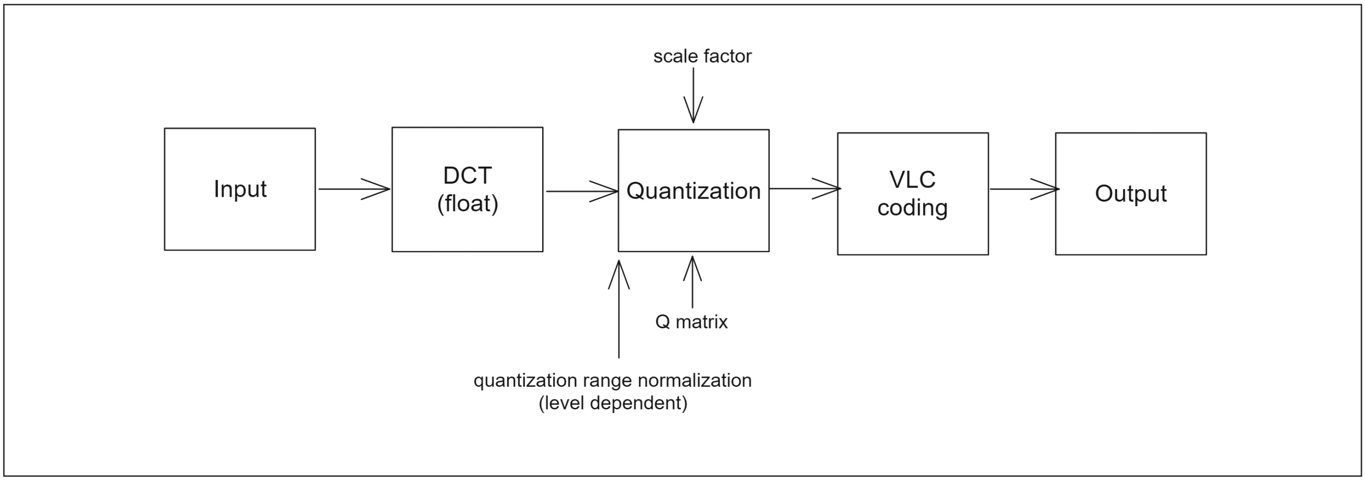


FIGURE 7. Block diagram of the processing that occurs during DNx encoding.

1545-0279/24©2024SMPTE

Part 4: Q Matrix Flexibility

The 8-bit issue can be addressed if you allow (optionally) a minor departure from the standard:

The values of the Q matrix, which is applied during the quantization stage, are not bit-depth dependent. The difference between the different Q matrices is primarily in how aggressively they eliminate the high-frequency DCT coefficients.

What differentiates the 444 level from the sub-sampled levels is that it uses the same Q matrix for all three components (4:2:2 and 4:2:0 levels use more robust quantization settings in the chroma channels), thus balancing the influence of the three components more equally in the rate controller.

To achieve an optimal result for 8-bit input, an implementation can choose the following approach:

During the quantization stage, normalize to the 8-bit design bit depth range (as defined in Amendment 1) and then use the 444-level Q matrix for quantization. Depending on the implementation, the bitrate controller will not care about this difference and will happily proceed. However, it will now start with a smaller footprint at the input to the quantization stage, allowing it to use less quantization pressure to achieve the desired target bitrate. Stated differently, it will lead to an uptick in quality compared to the 444 level.

For a software implementation, this is a trivial modification. However, this goes along with the provision that you can no longer expose the bitstream as a VC-3 compliant bitstream, as this is breaking standards compliance.

Part 5: RDD 50 DNxPacked Coding

RDD 50[5] (Avid DNxUncompressed) consists of two parts:

In Part 1, a simple, generic, and extensible, chunk-based bitstream super-format is introduced, which applies to all kinds of I-frame codecs. It allows flexibly storing image information in either single channels or as arbitrary combinations of channels. What is stored is described generically, while how it is stored relies on FourCC identifiers for the *chunks*. The format specifically provides generic guidelines for bitstream consumers to ignore/preserve chunks for which they do not know how to process the FourCC mapping of the *chunk*, and proceed with those chunks they can handle.

Part 1 of RDD 50, known as the DNxPacked bitstream format, does not specify any FourCCs. But it completely describes the mapping into the MXF container.

Part 2, DNxUncompressed, specifies a wide range of (lossless) channel codecs and their storage/packing format.

The generic nature of Part 1 allows adding new coding schemas to the bitstream by introducing new FourCCs without having to publish a complete codec specification, as long as they are I-frame formats. All it takes to embed a VC-3 coded frame into a DNxPacked bitstream is to define a new FourCC. Consumers who do not know this bitstream will skip it without having to interpret it, but the generic information will tell them enough to let them know what they are skipping and then make an informed decision if it makes sense to continue.

The resulting bitstream can be shared in MXF form without the risk of being confused with standard VC-3 bitstreams.

While this may look like a simple cheat schema, there is another very good reason always to encode such bitstreams this way, even if the compliant 444 level is used. VC-3 represents partially unfinished business in an irrelevant area for normal encodes, but tremendously important for graphics: Alpha.

For compatibility reasons, the bit depth of alpha in VC-3 is constrained to the same bit depth as the filler. This means that 10- and 12-bit input are constrained to using 10- and 12-bit alpha if the RLE compression option is chosen, as in this case, the "quantization range normalization" cannot be applied to the alpha.

While 10- and 12-bit alpha is not completely unheard of, most applications will use either 8- or 16-bit alpha. VC-3 also can neither support RLE compression for the 8-bit levels (size constraints) nor achieve 16-bit alpha resolution for the two higher quality levels.

This shortcoming is addressed by DnxPacked, which allows each channel to be handled independently. One of the channel codecs it supports by default is a variant of the RLE coding used by the lossless VC-3 alpha channel—but without the filler bit depth constraint of VC-3.

So, by putting the VC-3 compressed filler into a RDD 50-compliant container, this filler can be paired with an alpha channel of completely independent bit depth.

The importance of this combination goes well beyond what is known as "alpha blending." The alpha channel is a "general purpose" channel without any default operation/meaning attached to it. It is frequently used to store depth (3D) or velocity (particle systems) information, which relies on lossless and precise storage. So, the combination of VC-3 with the alpha channel handling of RDD 50 establishes a whole new use case for this codec, which goes well beyond that of traditional graphics applications.

Quality Considerations

After figuring out how to encode this, the obvious question looming is: Isn't this approach completely ruining the quality?

The surprising answer is, In some respects yes, but in practical use definitely no.

The most widely used Avid DNx level in 8-bit broadcast production is the standard quality (SQ) level at around 140 Mbits/s (4:2:2 at HD resolution). **Figure 8** shows the result when the RGB-based graphic is compressed to the same bitrate using the 444-level compression method. This certainly meets all the visual expectations and is at approximately a 13:1 compression ratio.

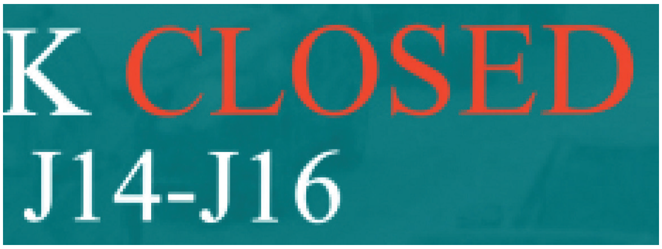


FIGURE 8. Result when running the 444 level at 8-bit design bit depth at the SQ level's target bitrate.

1545-0279/24©2024SMPTE

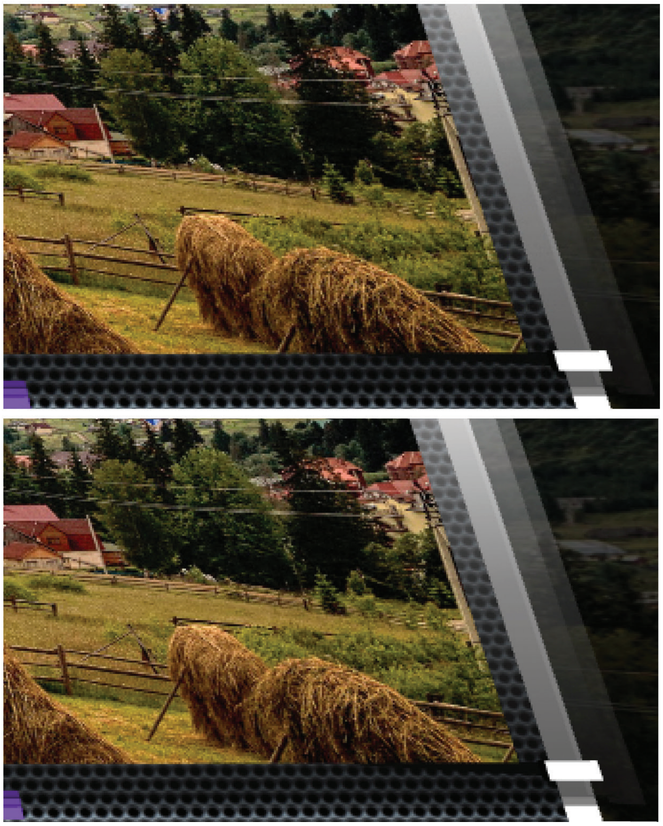


FIGURE 9. Comparison (500% zoom) of uncompressed original (top) and 444 SQ compressed (bottom).

This is further supported by looking at a detail of the "OPENER" example used before (**Fig. 9**). Not only do the graphics appear clear and well defined, as expected. The Live texture in the middle also does not show any visible artifacts when compared with the original.

Conclusion

Exploiting more exotic combinations of what is permissible by the VC-3 standard with minimal effort allows us to provide a solution for a graphics codec with a wide range of tunable target bitrates without defining an entirely new codec.

- Using VBR mode allows the use of much lower bitrates than targeted by the CBR mode.
- The generic rate controller required by VC-3's resolution independence design allows targeting such lower bitrates deterministically and without changes to the rate controller.
- The delayed bit depth limiting avoids visibly reducing the input quality, allowing quantization only in DCT space, where the quantization effects of one value will be distributed over 64 pixels after the IDCT.
- The level/bit depth independence of the Q matrix allows applying the 444 quantization matrix even at 8-bit design bit depth levels.
- Lastly, combining the compression with the chunk format of RDD 50 allows the creation of an exchangeable graphics format with independent alpha control.

While initially aimed at graphics, which starts with a noise advantage compared to real-life imagery, the results indicate that, counter to intuition, using an RGB/4:4:4-based DCT compression does have significant advantages when visual quality is more important than technical quality. This specifically applies to offline and proxy workflows, where using RGB as the native compression mode may also simplify viewing/browsing of the content on cheap and low-spec computer hardware (both CPU and monitors).

While the reference example used broadcast graphics, this solution and codec can be applied to graphics interop workflows involving virtual productions/volumes, VFX exchange, augmented graphics, or motion graphics. Generally, it can be applied wherever visual quality is deemed more important than technical quality.

A special case is employing the lossless alpha channel. The importance of a high-resolution and lossless alpha channel cannot be overstressed, as the alpha channel is a general-purpose channel that is not exclusively used for transparency blending. An important application of the alpha channel is to store 3D pixel depth information (z-depth). The combination of Avid DNx with RDD 50 opens the door to a much broader application of this codec than just broadcast graphics.

Acknowledgments

The author would like to thank Andrzej Wojdala and Martin Appleton, both members of the Avid Technology Server and Graphics group, for incentivizing the idea and providing invaluable reality checks.

References

1. SMPTE, ST 2019-1:2016, "VC-3 Picture Compression and Data Stream Format."
2. SMPTE, ST 2019-1: 2016, Amd 1: 2023, "VC-3 Picture Compression and Data Stream Format – Amendment 1."
3. SMPTE, RP 2019-2: 2016, "VC-3 Decoder and Bitstream Conformance."
4. SMPTE, RDD 36: 2015, "Apple ProRes Bitstream Syntax and Decoding Process."
5. SMPTE, RDD 50: 2019, "Avid DNxUncompressed – Packing Definition and Mapping into the MXF Generic Container."

Resources

Avid DNxGX Exporter for Adobe Media Encoder and Premiere Download (free)
Avid-DNxGX-Exporter-for-Adobe-Media-Encoder-Download
MediaCentral | Panel for Adobe Premiere Pro v2022.3 ReadMe
MCCUX_AdobePanel_2022_3_ReadMe.pdf

About the Author

Markus Weber studied physics and received a MSc in solid state physics in 1990 and a PhD in nuclear solid state/theoretical physics in 1992 from the Technical University of Munich. He is currently an architect in the Office of the CTO at Avid, where he is responsible for formats, codecs and engines. In his current role, he authored, among other things, the SMPTE ST 2019-1 (VC-3) standard, as well as RDD 50.

Presented at the 2023 Media Technology Summit. Hollywood, CA, 16-19 October 2023.

DOI: 10.5594/JMI.2024/YQTR8645
Date of publication: 24 January 2024

TECHNICAL PAPER

KEYWORDS FILM GRAIN // VIDEO COMPRESSION // CODING EFFICIENCY // HEVC // AOM // AV1 // H.265

Returning Greatness to Film Grain:

Introducing AV1-Compatible Film Grain Modeling for Existing HEVC-Based Video Codecs

By Dan Grois, Alex Giladi, Thomas Guionnet, Thomas Burnichon, Nikolay Tverdokhleb, and Mickael Raulet

Film grain applications started gaining a lot of popularity due to an increasing demand for natural appearance. To improve a video coding gain, the film grain is removed prior to encoding, and then added back after decoding at the post-processing stage, as is currently implemented in the AOM AV1 coding scheme.

1545-0279/24©2024SMPTE

Abstract

Film grain is spatially random in nature, while its physical size can also vary. In addition, film grain is independent in a temporal domain, thereby making it inherently difficult to compress. Therefore, to improve a video coding gain, the film grain is removed prior to encoding, and then added back after decoding at the post-processing stage, as is currently implemented in the Alliance for Open Media (AOM) AV1 coding scheme. In this work, the AV1-compatible film grain modeling for the High-Efficiency Video Coding (HEVC)-based video codecs has been carried out to utilize the existing AV1 film grain post-processing support efficiently. This is done by providing estimated film grain parameters within the International Telecommunication Union-Telecommunication (ITU-T) Recommendation T.35 (ITU-T T.35) Supplemental Enhancement Information (SEI) message. According to extensive experimental results conducted on popular cinematic content, significant bitrate savings are achieved.

Video applications continue to gain a lot of traction and have an enormous demand. As a result, bandwidth requirements continue to rise, mainly due to an increasing demand for ultra-high-definition (UltraHD) resolution displays (note that “UltraHD” refers to the 3840 x 2160 resolution (UHDTV-1/4K) in terms of luma samples). Since a typical bit rate for UHDTV-1 video is between 15 and 18 Mbits/s [1,2] it is generally more than twice the high-definition (HD) video bitrate and a factor of nine times the standard-definition (SD) video bitrate; as a result, there is currently a strong need to decrease video transmission bitrate substantially without reducing the visual presentation quality.[3]

During the last couple of years, artificially generated content has become very popular due to the technological breakthrough of Generative AI technologies. As a result, film grain applications started gaining a lot of popularity due to an increasing demand for natural appearance. Film grain is a product of the physical characteristics of analog film that refer to light-sensitive silver halide crystals (i.e., bromide, chlorine, iodine). These crystals vary in size due to the film’s exposure and development process, resulting in a random pattern of grain on the film. In addition, it should be noted that film grain differs in many aspects from digital noise (being visible after amplification) caused by electrons due to their thermal behavior in the sensor/electronics: for example, the size of grains varies depending on the film sensitivity (i.e., the more sensitive the film, larger the grains are), while the digital noise size is equal to the size of a pixel. On the other hand, in digital videos the film grain is added on purpose, mainly to provide desired artistic intent and natural look, or to mask coding artifacts due to compression.

Therefore, there is a need to preserve film grain while keeping the video transmission bitrate at a reasonable level, especially for the UltraHD video content.

Background

Main efforts to achieve the bitrate savings for UltraHD video content started in 2010, when ITU-T Video Coding Experts Group (VCEG) and the ISO/IEC Moving Pictures Expert Group (MPEG) established a Joint Collaborative Team on Video Coding (JCT-VC) to work on the High-Efficiency Video Coding (HEVC) standard.[3,4]

Then, after about two and a half years of intensive development, the first version of the HEVC specification was approved by ITU-T as Rec. H.265 and by ISO/IEC as MPEG-H, Part 2,[4] in 2013. **Figure 1** illustrates a block diagram of the H.265/MPEG-HEVC encoder—it should be noted that more detailed information about HEVC can be found at[5] and.[6] The HEVC second version contains the Range Extensions (RExt) as well as the Scalable and Multiview Extensions (SHVC and MV-HEVC, respectively), and it was accomplished in 2014.[7] Further, the third and fourth HEVC versions contain the 3D Video Coding Extensions (3D-HEVC)[8] and the Screen Content Coding Extension (HEVC-SCC),[9] and they were approved in 2015 and 2016, respectively.[10]

While the joint video coding standardization activities of ITU-T and ISO/IEC organizations rely on an open and collaborative process driven by its active members, several companies individually developed their video coding formats.[11,12,13] Technologically, these proprietary video coding formats were often designed around variants of the state-of-the-art coding algorithms supplemented with the companies’ proprietary technologies. For example, the VP8 codec[14] developed privately by On2 Technologies Inc., a company acquired by Google Inc. in 2009, is one such pro-

1545-0279/24©2024SMPTE

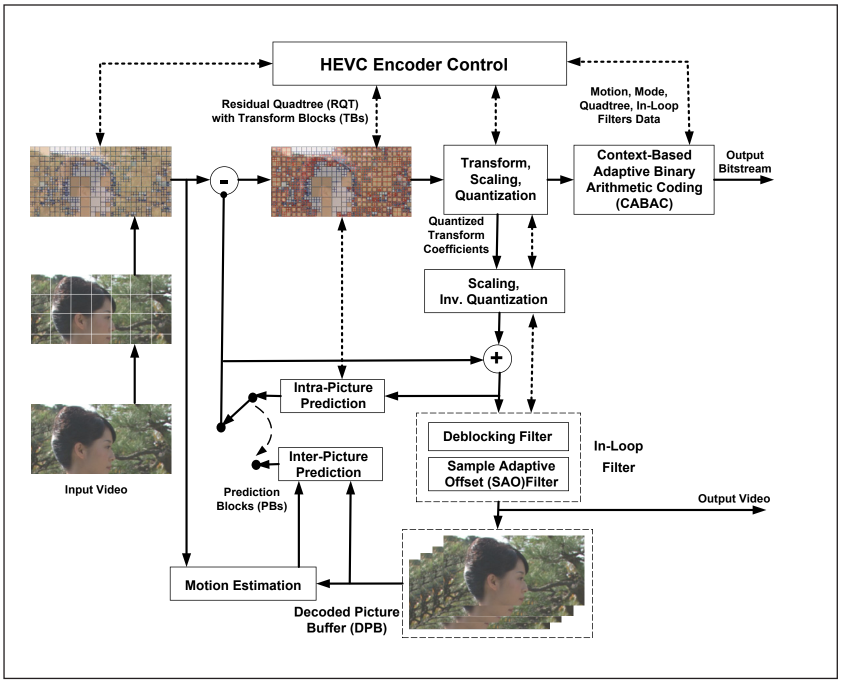

HEVC Encoder Control
Residual Quadtree (RQT) with Transform Blocks (TBs)
Motion, Mode, Quadtree, In-Loop Filters Data
Transform, Scaling, Quantization
Context-Based Adaptive Binary Arithmetic Coding (CABAC)
Output Bitstream
-
Quantized Transform Coefficients
Scaling, Inv. Quantization
+
Intra-Picture Prediction
Inter-Picture Prediction
Deblocking Filter
Sample Adaptive Offset (SAO)Filter
In-Loop Filter
Input Video
Prediction Blocks (PBs)
Output Video
Motion Estimation
Decoded Picture Buffer (DPB)

FIGURE 1. The schematic block diagram of the H.265/MPEG-HEVC encoder.

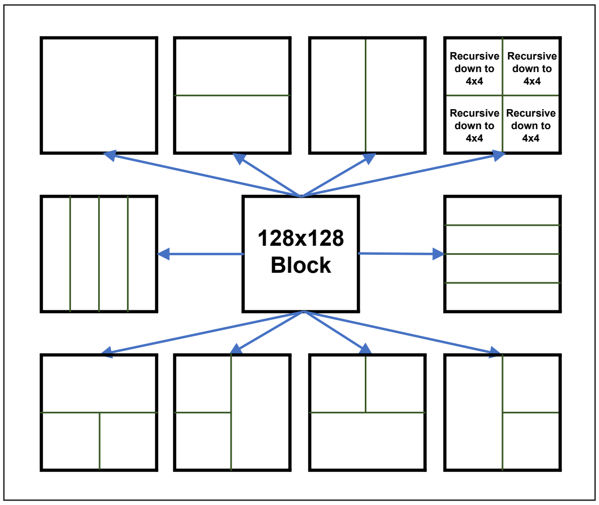


FIGURE 2. The schematic illustration of the AV1 codec's partitioning structure.

prietary video codec. In turn, the development of the VP9 codec,[15] as the successor of VP8, started about two years after the acquisition of On2 Technologies Inc. and was finalized in 2013.[16] Performance comparisons of VP9 with H.264/MPEG-AVC and H.265/MPEG-HEVC based on objective assessments were presented in detail.[11,12,13] In addition, subjective assessments of HEVC and VP9 can be found in Ref. 17.

About two years after VP9 was finalized, i.e., in 2015, the Alliance for Open Media (AOM) was formed by its founding members Amazon, Google, Microsoft, Cisco, Intel, Mozilla, and Netflix with the objective to work towards next-generation media formats in general and with a particular short-term focus on the development of a video-coding scheme.[13,18,19] The work on the new video-coding scheme in AOM started by using the source code of the above-mentioned VP9 codec, further adding advanced tools. Later, in April 2016, AOM released a baseline version of the developed video-coding scheme, which received the name AV1. Similarly, to HEVC, one of the main AV1 features is its partitioning structure, which is schematically presented in **Fig. 2**.

The final version of the first edition of AV1 was released in 2018, claiming to provide a significant coding-efficiency gain over the then-current state-of-the-art video codecs[16] and introducing film grain modeling, which was processed as follows (**Fig. 3**). First, the input video is denoised, and then the corresponding film grain parameters are estimated by analyzing the difference between the original input video and the denoised one. Then, the denoised video is encoded, and the resultative bitstream is provided to the decoder along with the estimated film grain parameters. In turn, the film grain is synthesized at the decoder end based on the received film grain parameters, and then the synthesized film grain is added to the reconstructed video.

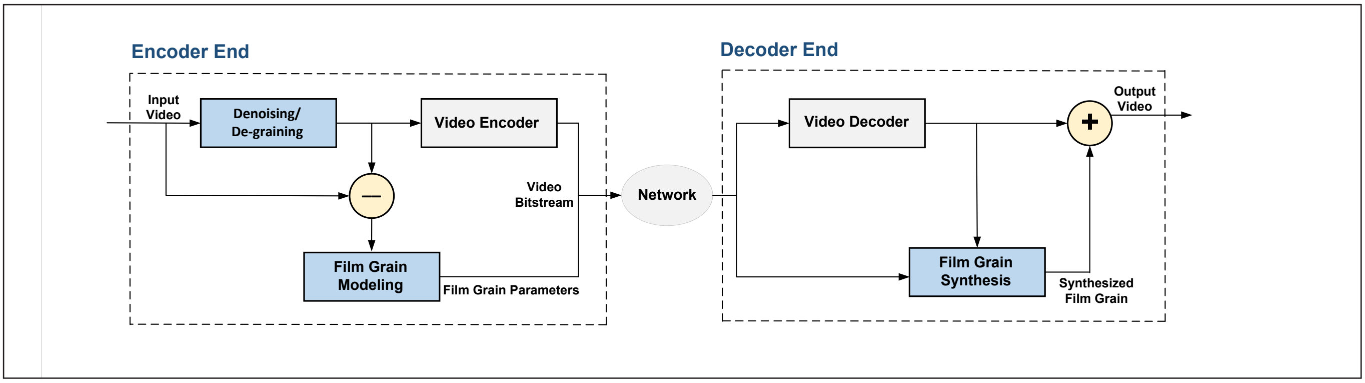


FIGURE 3. A schematic block diagram of the film grain processing within the coding loop.

1545-0279/24©2024SMPTE

The film grain modeling process in AV1 has been defined as *normative*, thereby making all AV1-compatible decoders to support and allowing the addition of the synthesized film grain during the post-processing step, i.e., prior to the display. It should be noted that HEVC also allows the film grain modeling process/synthesis by using metadata signaled within the film grain Supplemental Enhancement Information (SEI) message. However, in HEVC, it is not defined as mandatory, so, the AV1-based film grain processing is more widely deployed to date.

In this work, the AV1-compatible film grain modeling for the H.265/MPEG-HEVC-based video codecs has been carried out to utilize the existing AV1 film grain post-processing support efficiently. This is done by providing the estimated film grain parameters within the ITU-T Recommendation T.35 (ITU-T T.35) SEI message,[20] where the ITU-T T.35 header allows to indicate that the AV1 film grain parameters are provided within its registered AV1 Film Grain Specification (AFGS1) metadata payload to the decoder end[21]. Generally, the AV1 film grain synthesis parameters, which are specified within the AV1 specification,[19] are similar to the metadata payload mentioned earlier, including the frame-level film grain synthesis algorithms and frame-level parameters, which are similar to the frame-level parameters of the AV1 film grain synthesis process. However, there is a difference between the AFGS1 and AV1 film grain, which relates to the film grain parameters referencing, i.e., to how the film grain parameters are signaled.

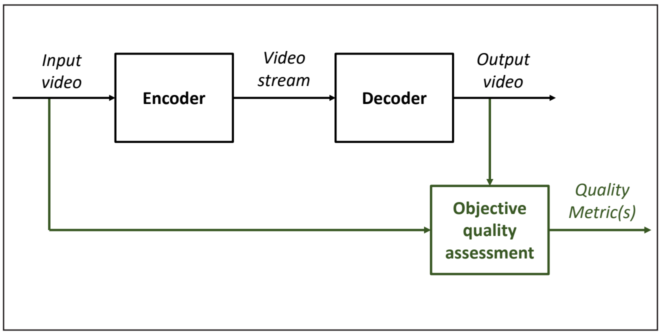


FIGURE 4. Video encoder objective evaluation.

The remainder of this paper is organized as follows. The AOM film grain signaling, as described within the AFGS1, is briefly discussed in the next section. Then, the selected representative encoder is introduced, followed by a description of the test methodology and evaluation setup. After that, detailed experiments that have been carried out on popular cinematic content are provided, and finally, conclusions are outlined in the last section of this work.

AOM Film Grain Signaling

For the past few years, AOM has been working on the AV1 film grain synthesis specification, which is abbreviated as AFGS1.[21] According to the specification, the estimated film

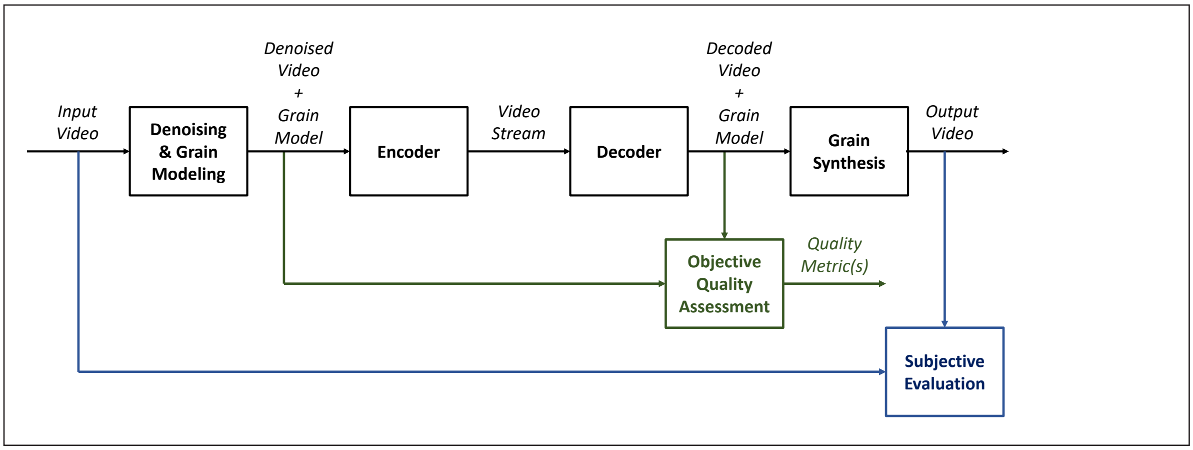

Input Video
Denoising & Grain Modeling
Denoised Video + Grain Model
Encoder
Video Stream
Decoder
Decoded Video + Grain Model
Grain Synthesis
Output Video
Objective Quality Assessment
Quality Metric(s)
Subjective Evaluation

FIGURE 5. Video encoder and film grain synthesis evaluation.

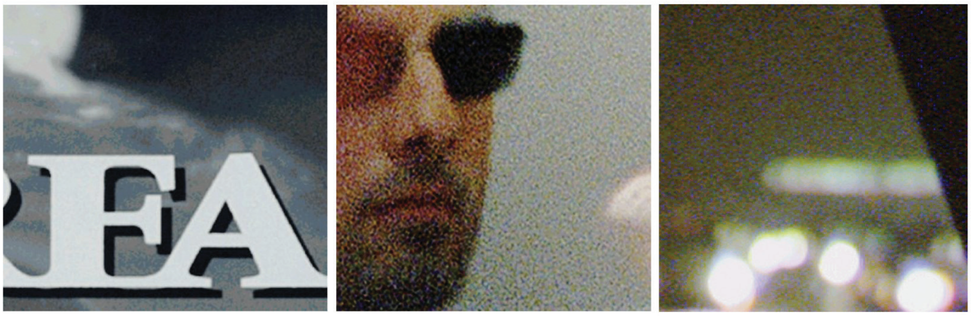


FIGURE 6. "Collateral" input video samples.

grain parameters are provided within the ITU-T Rec. T.35 (ITU-T T.35) SEI message. Specifically, a syntax element titled *itu_t_t35_terminal_provider_oriented_code* is used within the ITU-T T.35 header for indicating the signaling of AV1 film grain parameters within the ITU-T T.35 registered metadata payload.

For keeping the film grain in AFGS1 as codec-agnostic, no concept of reference frames is used. A specific flag (*apply_grain*) is utilized for indicating the application of film grain synthesis to a current frame. At the same time, signalling of "1" means that the film grain is applied to the current frame and when that happens, an additional set of parameters is signaled: *grain_seed* for initializing the pseudo-random numbers generator; *film_grain_param_set_idx* parameter for allowing to differentiate between up to eight separate film grain models (e.g., referring to different output resolutions) as well as *update_grain* parameter. In turn, when *update_grain* is "1", then it means that a new set of film grain parameters should be parsed from the metadata payload to be applied to the current frame, further being associated with specific *film_grain_param_set_idx*. On the other hand, when *update_grain* is 0, then the film grain parameters are read from the memory associated with the *film_grain_param_set_idx* to be further applied to the current frame.

It should be noted that AFGS1 can be integrated with many video codecs that support ITU-T T.35 metadata. Naturally, the AFGS1 film grain should be supported throughout the post-processing process to ensure that every component works together, providing significant benefits, for example, when transcoding video between different formats or encoding video in multiple formats. The reader can find more detailed information in Ref. 21.

Selected Encoder Implementation: The Titan Encoder

For carrying out an evaluation of the AV1-compatible film grain modeling for the H.265/MPEG-HEVC-based video codecs, the Ateme Titan proprietary video encoder was selected.[22] Titan codec is considered to be a professional contribution encoder, which is capable of encoding video sequences in both HEVC and AV1. For removing the film grain, the Titan encoder incorporates adaptive MCTF-based denoising, complemented by frequency domain spatial filtering for temporally unpredictable areas and grain model analysis.

The Titan encoder (version 36.0.0.0 of 7 August 2023)

1545-0279/24©2024SMPTE

has been selected for both the AV1-based and the HEVC-based encoding—to date, it is the only available encoder that allows the AFGS1 film grain signaling by using ITU-T T.35 registered metadata for the HEVC-based encoding.[22]

Test Methodology and Evaluation Setup

Figure 4 depicts a typical video encoder evaluation technique in which an objective quality assessment is performed by comparing an output video to the input video while applying objective quality metrics. After completing several encodings, a Bjøntegaard-Delta Bit Rate (BD-BR) measurement method[23] is typically used for the Rate-Distortion (R-D) performance assessment to calculate average bitrate differences between R-D curves for the same distortion, e.g., for the same $\mathrm{PSNR}_{\mathrm{YUV}}$ values.[24] It should be noted that while conventional objective quality assessment methods are not perfect, they have proven useful.

On the other hand, film grain synthesis, the approach mentioned, is not useful: the grain synthesis process generates a pseudo-random noise with the same statistical characteristics as the source video grain. Therefore, it retains the grain look as much as possible but has no fidelity to the source. In other words, it is a different grain (varying by size, density, and color) with substantially the same subjective appearance/look. As a result, any objective metric comparing the output video to the input video will determine a significant degradation in visual quality because the grain pattern is different. The corresponding encoding/decoding process is schematically illustrated in **Fig. 5**.

Alternatively, before adding the film grain at the post-processing step, the decoded video can be compared to the input video. However, also in this case, an objective metric comparing the decoded video to the input video will provide non-useful results since the decoded video is a denoised version of the input video, which has been encoded. Therefore, the only valid objective comparison should be performed between the decoded output video (before adding a film grain prior to display) and the denoised input video (after performing the film grain denoising/removal). The final output video can then be assessed subjectively.

In this work, the experimental evaluation of the film grain synthesis (FGS) is decomposed into two parts: coding efficiency and grain rendering. First, the coding efficiency is evaluated objectively, and then validated subjectively. Second, the grain rendering is visually inspected. For that, the expert viewing has been conducted – see Section: *Viewing Session Setup*.

The process of film grain parameters extraction for the purpose of the film grain synthesis is described in the next section.

Film Grain Parameters Extraction

The Titan encoder,[22] which was selected to run experi-

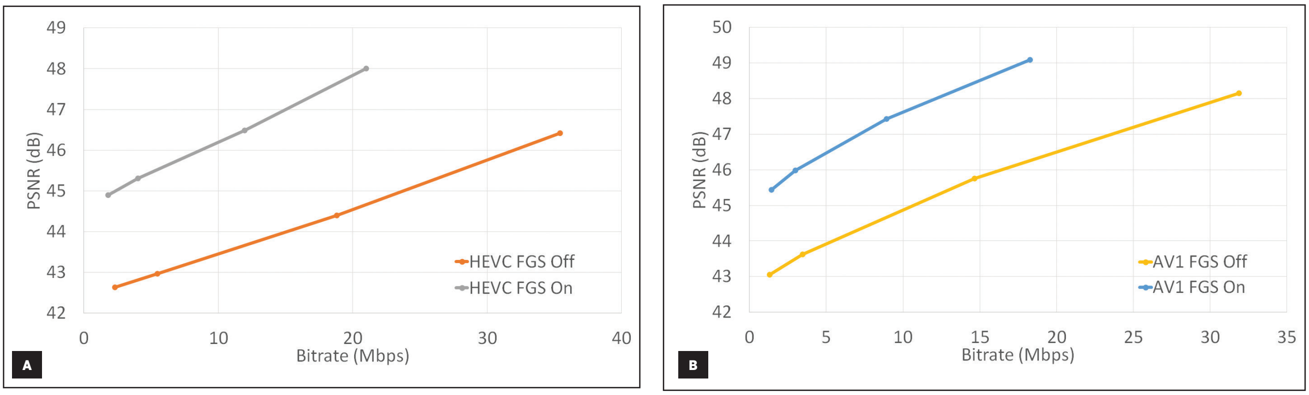


FIGURE 7. "Collateral" R-D curves for CQ encodings with/without FGS: (a) HEVC, (b) AV1.

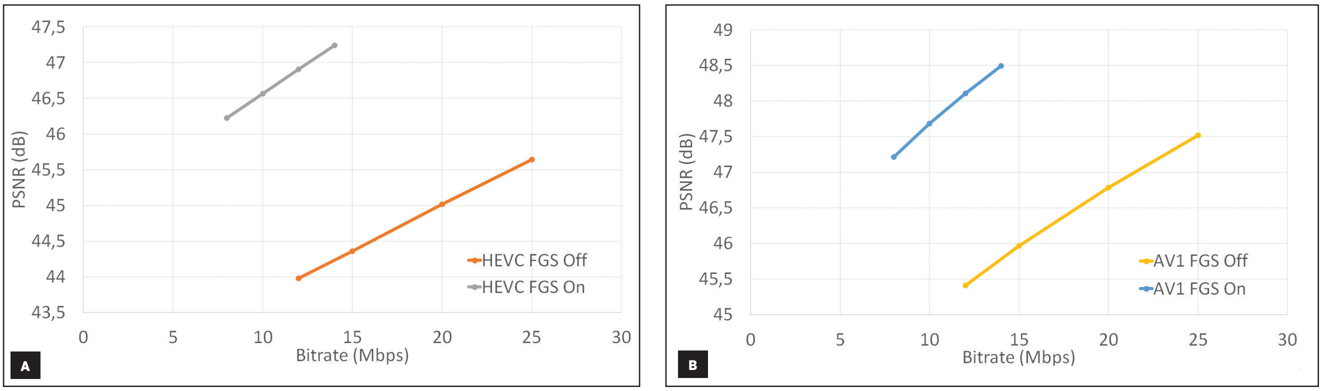


FIGURE 8. "Collateral" R-D curves for CBR encodings with/without FGS: (a) HEVC, (b) AV1.

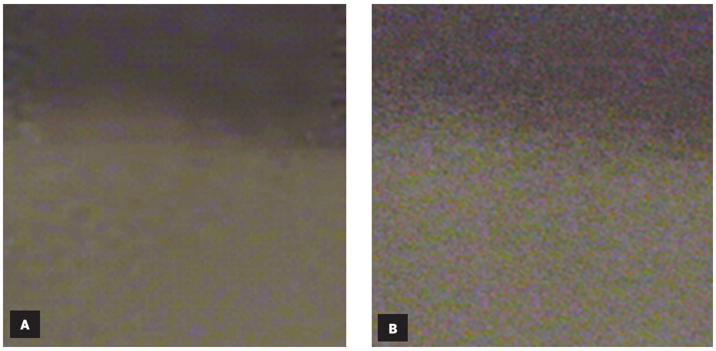


FIGURE 9. "Collateral" AV1 encoding at 12 Mbits/s, (a) FGS Off, (b) FGS On

ments in this work, must perform extraction of film grain parameters before denoising/encoding of the film-grained video content for further signaling these parameters by using the ITU-T T.35 SEI message.[20] For that the Titan encoder performs the following operations:

- **Luma and chroma range segmentation:** As specified in AV1 specification[19] and in AFGS1,[21] the luminance and chrominance ranges can be split into bands, each band having its own scaling or grain "strength." This step is systematically applied as the grain usually exhibit a variation depending at least on luminance. In addition, there are two design criteria. First, the range must be split into small enough bands to ensure fine scaling adaptation. Second, each band must contain enough data to be statistically significant. A non-uniform splitting of the range is preferred. Starting from the maximum possible number of bands, the Lloyd-Max algorithm[25] is applied with specific initializations for luminance and chrominance channels. Then, the adequacy of the design criteria is checked. If necessary, the process is repeated with fewer possible bands.

- **Identification of reference picture blocks:** For each band, a subset of blocks is selected for grain parameters estimation. These blocks are set either on the source image or the residual image after motion compensation (as performed by the MCTF filter). Generally speaking, the blocks with the lowest energy are supposed to contain only film grain. As such, they can be used as references for grain parameter estimation. In the spatial domain, it corresponds to uniform areas, while in the temporal domain, it corresponds to blocks with perfectly estimated motion. With the exclusion of specific areas, such as black stripes, the reference blocks are selected as a combination of source and residual blocks with the lowest energy.

- **Grain variance estimation:** For each band, the variance of the grain is estimated in the reference blocks, and the scaling values are set accordingly. It should be noted that the overall distribution of the scalings is checked, and some values may be corrected if they are inconsistent with the neighboring bands.

- **Auto-Regressive (AR) coefficient estimation:** Similarly, the reference blocks are used for AR coefficient estimation. However, as the AR coefficients are defined once for all the ranges, the reference blocks from the highest scaling bands are gathered. The Yule-Walker method is considered in Ref. 26. However, the convergence is not guaranteed, and the coefficients may need

TABLE 1. Configuration settings of the Titan encoder for the HEVC-based encoding.

HEVC Titan encoder configuration			
Coding options	**Chosen parameters**	**Coding options (cont.)**	**Chosen parameters (cont.)**
Encoder Version	36.0.0.0, Aug. 7, 2023	Rate Control	Enabled
Profile	Main 10	Intra Period	1 sec
Level	5.0	Coding Unit Size/Depth	64/3
GOP	32	Transform Unit Size (Min/Max)	4/32
Threads	8	Deblocking Filter	Enabled
Lookahead	23	Sample adaptive offset (SAO)	Enabled
Optimizations	AVX2	Asymmetric Motion Partitioning	Enabled

TABLE 2. Configuration settings of the Titan encoder for the AV1-based encoding.

AV1 Titan encoder configuration			
Coding options	**Chosen parameters**	**Coding options (cont.)**	**Chosen parameters (cont.)**
Encoder Version	36.0.0.0, Aug. 7, 2023	Rate Control	Enabled
Profile	Main	Intra Period	1 sec
GOP	32	SB size	64
Threads	8	Coding Unit Size/Depth	64/3
Lookahead	23	Transform Unit Depth	3
Optimizations	AVX2	Deblocking Filter	Enabled

1545-0279/24©2024SMPTE

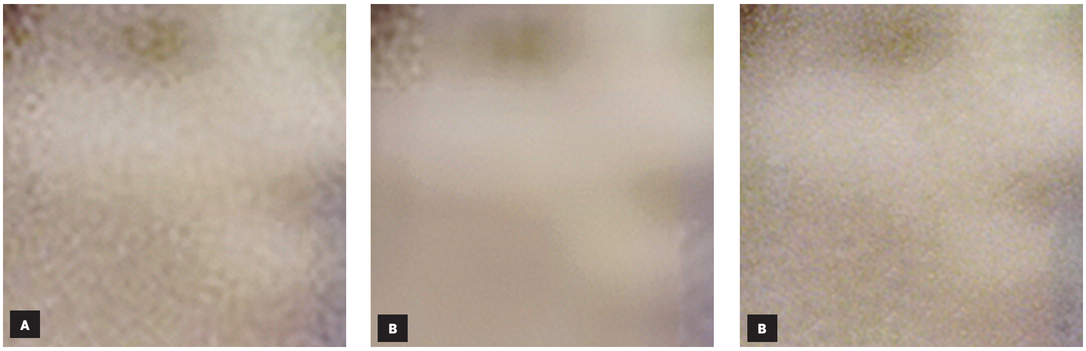


FIGURE 10. "Collateral" AV1 encoding, (a) FGS Off, 25 Mbits/s, (b) FGS Off 12 Mbps, (c) FGS On 12 Mbps.

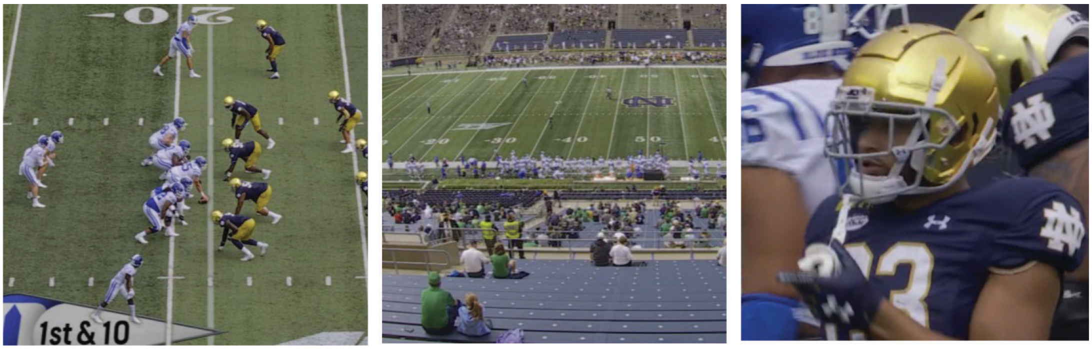


FIGURE 11. "Notre Dame Football" input video samples.

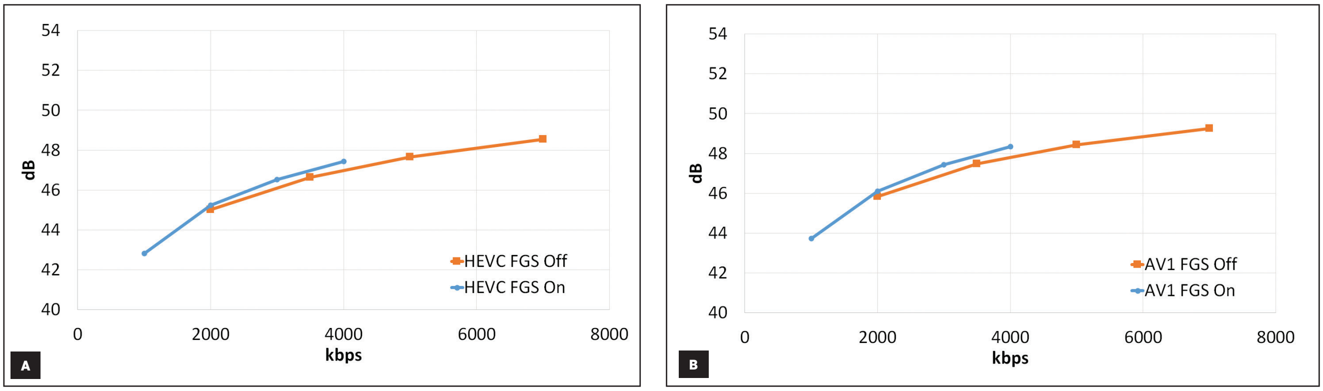


FIGURE 12. "Notre Dame Football" R-D curves for CBR encodings with/without FGS: (a) HEVC, (b) AV1.

to be adjusted, as the standard constrains the range of the AR coefficients. Therefore, the ridge regression[27] is preferred. As the lambda parameter controls the norm of the AR coefficients, it is possible to control both convergence and coefficients range. However, very coarse grain may be underestimated because of these constraints.

- **Seed selection:** The seed for the pseudo-random generator is selected. It has been observed that some unpleasing repetitive patterns appear due to the replication of a small patch of grain. At this step, a seed is selected among a subset known for being less prone to the apparition of repetitive patterns.

The Titan encoder configuration settings (both for HEVC-based and AV1-based encoding) are discussed in detail in the next section.

Titan Encoder Configuration: HEVC Encoding
The detailed configuration settings of the Titan encoder for performing the HEVC-based encoding are presented in **Table 1**. The Random Access (RA) configuration was selected, because it provides the most significant coding gains.

Titan Encoder Configuration: AV1 Encoding
To ensure a fair performance evaluation, the authors tried to provide similar configuration settings for the AV1-base encoding as well. **Table 2** summarizes the encoder configuration settings.

The expert viewing session setup is provided below.

1545-0279/24©2024SMPTE

Viewing Session Setup

The viewing was conducted according to adjectival categorical judgment described in Annex 4, Part 2 of the "Stimulus-comparison methods," BT.500-14,[28] with expert viewers. The viewing was conducted on UHDTV-1 65 in. diagonal TV sets, which use OLED technology, in dedicated rooms with dim lighting.

The post-processing functions of all TV sets have been deactivated. The viewing distance was 1.6 times the screen height. The voting procedure was "A"-"B"-"A"-"B," where "A" is an anchor sequence and "B" is a test sequence.

Also, the viewers rated "B" versus "A" as one of: "much worse," "slightly worse," "same," "slightly better," "much better."

Bjøntegaard-Delta Bit-Rate Measurements

Similarly, to the previous work,[2] the Bjøntegaard-Delta Bit Rate (BD-BR) measurement method[23] was used for the R-D performance assessment in order to calculate average bitrate differences between R-D curves for the same distortion, e.g., for the same PSNR_{YUV} values.[24] *It should be noted that negative BD-BR* values indicate actual bitrate savings.

The authors used R-D curves of the combined luma (Y) and chroma (U,V) components, while the combined PSNR_{YUV} values are calculated as a weighted sum of the PSNR values per each picture of each component, i.e., of PSNR_Y, PSNR_U, and PSNR_V.

$$\text{PSNR}_{YUV} = (6 \bullet \text{PSNR}_Y + \text{PSNR}_U + \text{PSNR}_V) / 8$$ [1]

As a result, using the combined PSNR_{YUV} and bitrate values as an input to the BD-BR measurement method enables determining a single average difference in bitrate that takes into account the reconstruction fidelity of both the luma and the two chroma components.[2,24]

Experimental Results and Brief Discussion

The experiments were conducted by using three test video sequences, the details of which are provided in **Table 3**.

TABLE 3. Test video sequences.

Sequence name	Collateral	Notre Dame Football	Blues Brothers
Resolution	3840 × 2160 (UHDTV-1)	1920 ×1080 (1080p)	1920×1080 (1080p)
Bit-Depth	10 bits	10 bits	10 bits
Duration	10 min	10 min	7 min 44 sec
Frame Rate	23.976 frames/sec	59.94 frames/sec	23.976 frames/sec
Color primaries	BT.709	BT.2020	BT.709

The "Collateral" video sequence contains several grain characteristics, including light grain overlayed on synthetic content, very strong/coarse film grain, and digital grain coming from night capture (as shown in **Fig. 6**).

The experiments have been carried out in the following way: First, the video sequence was encoded at the constant quality (CQ) mode, where a target encoding quality is predefined.

PERFORMING THE **FILM GRAIN SYNTHESIS** IS PROVEN TO BE THE BEST WAY OF CONVEYING FILM GRAIN AT LOW BITRATES, FOR THE **ULTRAHD** VIDEO CONTENT.

Similarly, to predefining a quantization parameter (QP), the denoised (de-grained) video requires fewer bits to be encoded at the same visual quality level as the original input video.

Then, in the second step, the video sequence was encoded in the constant bitrate (CBR) mode, with bitrates reflecting the CQ encodes observations. For each experiment, four rate-distortion points are generated, with the film grain synthesis turned ON and OFF (i.e., with and without adding film grain after the decoding and before display).

The CQ encoding results are presented in **Fig. 7**. As expected, the denoised video exhibits a dramatically better coding efficiency, with around 80% BD-BR gain for both AV1 and HEVC. Similar results have been obtained in CBR mode in **Fig. 8**.

Upon conducting the subjective assessment, it was found that the consistent benefit of the FGS is the grain rendering stability at low bitrates, such as 15 Mbits/s or lower for the UltraHD video content.

Regarding grain "look," the actual results depend on the characteristics of the video source. In the example of in **Fig. 9**, the film grain aspect is satisfactory. On the other hand, as illustrated in **Fig. 10**, some very coarse grain patterns may prove to be challenging. Once again, the grain is stable temporally and spatially due to performing the FGS, but the overall film grain appearance may differ from the source. One must note that in case of very coarse grain, the denoiser may struggle too and not fully remove the grain.

As illustrated in **Figs. 9 and 10**, film grain is either removed by the encoding process, or worse, unstable spatially and temporally. The grain can be also considered as "pumping"—i.e., appearing and disappearing depending on the frame position within the group of pictures (GOP) and depending on the luminance characteristics. Therefore, to achieve the best visual/subjective performance at

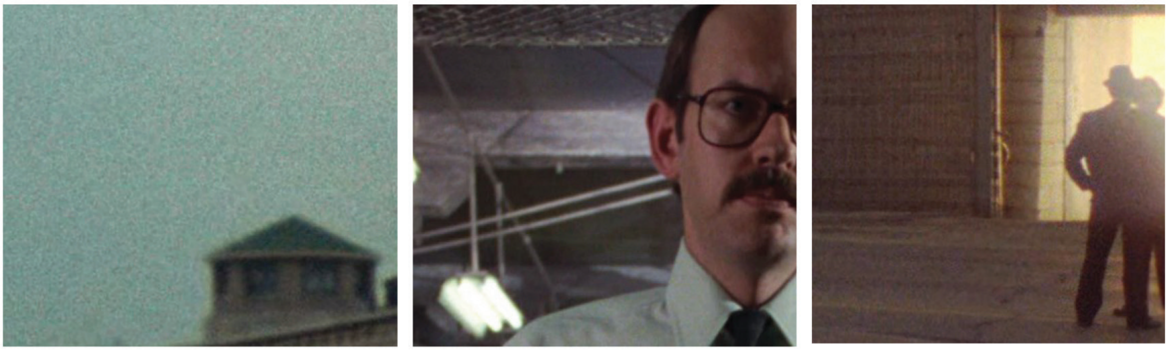


FIGURE 13. "Blues Brothers" input video samples.

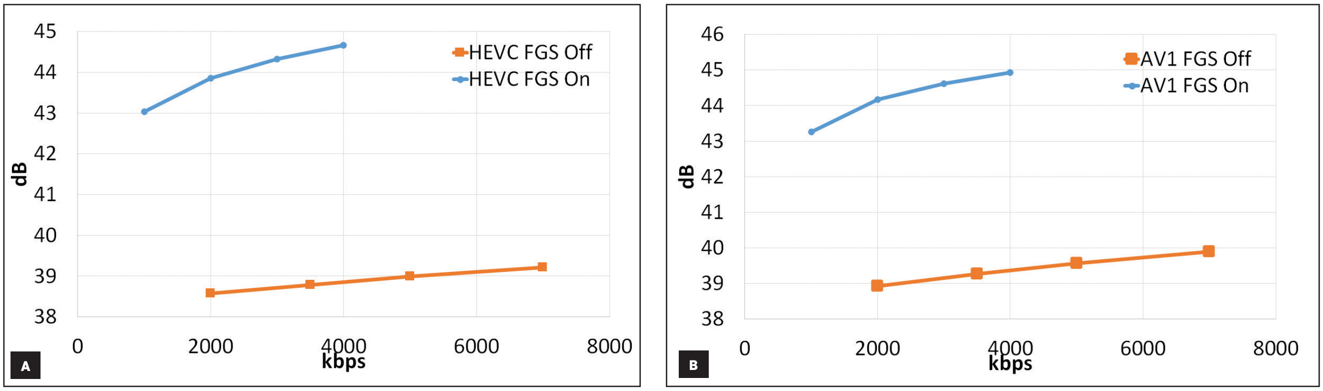


FIGURE 14. "Blues Brothers" R-D curves for CBR encodings with/without FGS: (a) HEVC, (b) AV1.

the decoder end, a mix of the residual grain and synthesized grain should be carried out.

According to the experimental results, by applying the FGS to the decoded video content, the overall bitrate reduction of ~40% is achieved for the "Collateral" video sequence, with satisfactory visual quality.

In addition to the "Collateral" video samples presented in **Fig. 6**, the "Notre Dame Football" video samples are presented in **Fig. 11**. As it is seen, the "Notre Dame Football" video sequence is much less grainy, and it includes many fast motion scenes.

The corresponding CBR encoding results are presented in **Fig. 12**. The BD-BR gain for the "Notre Dame Football" video sequence for both AV1 and HEVC-based encoding is more than 10%; however, it is much lower than for the "Collateral" video sequence due to much less grainy video source.

The experimental results for the "Notre Dame Football" video sequence clearly show that there are significant benefits for using the FGS technique even in the case of the very low film grain, since it still allows one to significantly reduce bitrate and increase coding gain.

In addition, the "Blues Brothers" video samples are presented in **Fig. 13**. The corresponding CBR encoding results of the "Blues Brothers" video sequence are presented in **Fig. 14**. As it is seen, the BD-BR gain for both AV1 and HEVC is very significant, and it is more than 90%.

Table 4 presents a summary of the BD-BR bit rate savings for the tested video sequences. It should be noted that negative BD-BR values indicate actual bitrate savings.

TABLE 4. A summary of the BD-BR bit rate savings.

Sequence	HEVC Bit Rate Savings FGS ON vs. FGS OFF	AV1 Bit Rate Savings FGS ON vs. FGS OFF
Collateral	-79.83%	-68.12%
Notre Dame Football	-10.02%	-12.27%
Blues Brothers	Gain of more than 90%	Gain of more than 90%

As shown in **Table 4**, the HEVC bitrate saving for video sequences with heavy film grain, such as "Collateral" and "Blues Brothers," are very significant and they are ~80% and more than 90%, respectively. On the other hand, for the "Notre Dame Football" video sequence that has very light film grain, the bitrate savings are ~10%, which is still very significant considering that the level of grain/noise is very low.

Conclusion

This work presents an efficient implementation of the existing AV1 film grain post-processing support by using AV1-compatible film grain modeling for the H.265/

MPEG-HEVC based video codecs. This task has been accomplished by providing estimated film grain parameters within the ITU-T Rec. T.35 (ITU-T T.35) SEI message. According to the extensive experimental results conducted on popular cinematic content, very significant bitrate savings are achieved for substantially the same subjective video presentation quality.

In this work, performing the film grain synthesis is proven to be the best way of conveying film grain at low bitrates, such as 15 Mbits/s or lower, for the UltraHD video content. The future work will include denoising and film grain modeling refinements.

References

1. D. Grois, and A. Giladi, "Perceptual quantization matrices for high dynamic range H.265/MPEG-HEVC video coding", *Proc. SPIE 11137, Applications of Digital Image Processing XLII, 111370O*, 2020.
2. D. Grois *et al.*, "Performance Comparison of Emerging EVC and VVC Video Coding Standards with HEVC and AV1," in SMPTE *Mot Imag. J. 130 (4):* 1-12, May 2021, doi: 10.5594/JMI.2021.3065442.
3. D. Grois, T. Nguyen, and D. Marpe, "Coding Efficiency Comparison of AV1/VP9, H.265/MPEG-HEVC, and H.264/MPEG-AVC Encoders," Picture Coding Symposium (PCS), Nuremberg, Germany, Dec. 2016.
4. International Telecommunication Union-Telecommunication (ITU-T), Recommendation H.265 (04/13), Series H: "Audiovisual and Multimedia Systems, Infrastructure of audiovisual services – Coding of Moving Video, High Efficiency Video Coding," Apr. 2013.
5. D. Grois, B. Bross, D. Marpe, and K. Sühring, "HEVC/H.265 Video Cding Standard: Part 1." [Online]. Available: https://www.youtube.com/watch?v=TLNkK5C1KN8&t=764s
6. D. Grois, B. Bross, D. Marpe, and K. Sühring, "HEVC/H.265 Video Coding Standard: Part 2." [Online]. Available: https://www.youtube.com/watch?v=V6a1AW5xyAw&t=5s
7. International Telecommunication Union-Telecommunication (ITU-T), Recommendation H.265 (10/14), Series H: "Audiovisual and Multimedia Systems, Infrastructure of audiovisual services – Coding of Moving Video, High Efficiency Video Coding," Oct. 2014.
8. International Teleommunication Union-Telecommunication (ITU-T), Recommendation H.265 (04/15), Series H: "Audiovisual and Multimedia Systems, Infrastructure of audiovisual services – Coding of Moving Video, High Efficiency Video Coding" Apr. 2015.
9. International Telecommunication Union-Telecommunication (ITU-T), Recommendation H.265 (12/16), Series H: "Audiovisual and Multimedia Systems, Infrastructure of audiovisual services – Coding of Moving Video, High Efficiency Video Coding," Dec. 2016.
10. D. Grois, T. Nguyen, and D. Marpe, "Performance comparison of AV1, JEM, VP9, and HEVC encoders," *Proc. SPIE 10396*, *Applications of Digital Image Processing XL, 103960L*, 2018.
11. D. Grois, D. Marpe, A. Mulayoff, B. Itzhaky, and O. Hadar, "Performance comparison of H.265/MPEG-HEVC, VP9, and H.264/MPEG-AVC encoders," *Proc Picture Coding Symposium (PCS)*, 2013, pp.394-397, 8-11 Dec. 2013.
12. D. Grois, D. Marpe, T. Nguyen, and O. Hadar, "Comparative Assessment of H.265/MPEG-HEVC, VP9, and H.264/MPEG-AVC Encoders for Low-Delay Video Applications", *Proc. SPIE Vol. 9217*, Sept. 2014.
13. D. Grois, T. Nguyen, and D. Marpe, "Coding Efficiency Comparison of AV1/VP9, H.265/MPEG-HEVC, and H.264/MPEG-AVC Encoders," *Proc. Picture Coding Symposium (PCS)*, pp. 1-5, Dec. 2016.
14. J. Bankoski, P. Wilkins, and X. Yaowu, "Technical overview of VP8, an open source video codec for the web," *Multimedia and Expo (ICME), 2011 IEEE International Conference on*, pp. 1,6, 11-15 Jul. 2011.
15. Chromium® open-source browser project, VP9 source code. [Online]. Available: http://git.chromium.org/gitweb/?p=webm/libvpx.git;a=tree;f=vp9;hb=aaf61dfbcab414bfacc3171501be17d191ff8506
16. Y. Chen et al. "An Overview of Coding Tools in AV1: The First Video Codec from the Alliance for Open Media," *APSIPA Trans. on Sig. and Inform. Process.*, Cambridge University Press, Feb. 2020.
17. M. Řeřábek, and T. Ebrahimi, "Comparison of compression efficiency between HEVC/H.265 and VP9 based on subjective assessments," *Proc. SPIE 9217, Applications of Dig. Image Process. XXXVII*, 92170U, 2014.
18. Alliance for Open Media, Press Release. [Online]. Available: http://aomedia.org
19. Git repositories on aomedia [Online]. Available: https://aomedia.googlesource.com/aom
20. ITU-T Recommendation T.35 Terminal Provider Codes. [Online]. Available: https://www.itu.int/en/ITU-T/inr/forms/Pages/t35.aspx
21. AOMedia film grain synthesis specification 1 (AFGS1). [Online]. Available: https://github.com/AOMediaCodec/afgs1-spec
22. Ateme Titan encoder. [Online]. Available: https://www.ateme.com/product-titan-software/
23. G. Bjøntegaard, "Calculation of Average PSNR Differences Between RD-Curves," ITU-T Q.6/SG16 VCEG 13th Meeting, Document VCEG-M33, Austin, TX, Apr. 2001.
24. J. Ohm, G.J. Sullivan, H. Schwarz, T.K. Tan, and T. Wiegand, "Comparison of the coding efficiency of video coding standards—including High Efficiency Video Coding (HEVC)," *IEEE Transactions on Circuits and Systems for Video Technology*, 22 (12): 1669-1684, Dec. 2012.
25. S.P. Lloyd, "Least squares quantization in PCM", *IEEE Trans. on Information Theory*, 28 (2): 129–137, 1982.
26. A. Norkin and N. Birkbeck, "Film Grain Synthesis for AV1 Video Codec," *2018 Data Compression Conference*, pp. 3-12, Snowbird, UT, 2018.
27. D.E. Hilt and D.W. Seegrist, "Ridge, a computer program for calculating ridge regression estimates," 1977.
28. International Telecommunication Union-Radiocummunication (ITU-R) Recommendation BT.500-14, "Methodologies for the subjective assessment of the quality of television images," Oct. 2019.

About the Authors

Dan Grois received a PhD from the Ben-Gurion University of the Negev (BGU), Israel. He founded several startup companies, and actively participates in the video standardization activities by contributing to the development of H.265/MPEG-HEVC and H.266/MPEG-VVC video coding standards.

Alex Giladi is a Comcast Fellow and an Emmy-winning technologist. He has been working on issues related to video content encoding and distribution since 2000. His current areas of interest are video encoding and adaptive streaming. He currently leads an advanced technologies group within Comcast.

Thomas Guionnet is a Fellow Research Engineer at Ateme, where he currently leads the Innovation team's research on artificial intelligence applied to video compression. Beyond his work for Ateme, he has also contributed to standardization process in multiple standards groups.

Thomas Burnichon is Ateme's director of technology and focuses on the video transcoding workflows for Live, video on demand and over-the-top streaming applications. He keeps in close contact with Content and Service Providers all over the world to identify best practices and help drive innovations.

Nikolay Tverdokhleb is a research engineer at Ateme, where he is conducting research on real-time and file-based video processing and encoding. He represents Ateme in MPEG with a focus on Joint Video Expert Team activities about standardizing the next-generation codec.

Mikael Raulet is chief technology officer at Ateme, where he drives research and innovation with various collaborative R&D projects. He represents Ateme in several standardization efforts. He is the author of numerous patents and more than 100 conference and journal scientific papers.

Presented at the 2023 Media Technology Summit. Hollywood, CA, 16-19 October 2023.

DOI: 10.5594/JMI.2024/JUNZ2729
Date of publication: 24 January 2024

Access to Self-Study Video Library

What’s New for Members in 2024!

Access to the entire SMPTE Independent Study Virtual Course Library

Education is more accessible too. Members can access self-study courses at no additional cost, so you can learn at your own pace! In addition, instructor-led classes are available to members at a reduced price. Continue your education with SMPTE!

EVOLVE with us
SMPTE

KEYWORDS IP-BASED WORKFLOWS // LOAD TESTING // SYSTEM LIMITATIONS // PREDICTABLE PERFORMANCE

Profiling the ST 2110 Network Traffic for Load Testing with Open-Source Tools

By Ievgen Kostiukevych, Thomas Kernen, Willem Vermost, and Pavlo Kondratenko

Abstract

The transition to IP-based and software workflows in broadcasting has increased demands on networks for consistent performance, especially with the adoption of the SMPTE ST 2110 standard and its unique timing, synchronization, and bandwidth requirements. Unlike traditional internet-like traffic profiles, ST 2110 systems exhibit distinct traffic characteristics, highlighting the need for standardized acceptance and load testing methods. This paper discusses profiling operational ST 2110 systems under load, creating a statistical model for media stream distribution based on typical use cases. It also explores using open-source tools for load testing, enabling media companies and broadcasters to simulate high-traffic scenarios, identify system limitations, and improve network scalability. This study advances the understanding of how network designs and switching platforms respond to high volumes of ST 2110-like traffic, offering a cost-effective, efficient, and scalable approach to network management in the broadcast industry.

The media production and broadcasting industry is experiencing what is arguably the most significant transformative shift since the introduction of the Serial Digital Interface (SDI).[1] Since then, a series of innovations have emerged, culminating in the first major milestone for media over Internet Protocol (IP), the AES67 standard,[2] and the ever-expanding SMPTE ST 2110 suite of standards.[3] This evolution aims to replace traditional broadcasting methods with agile, scalable, cost-effective media over IP solutions powered by advanced IP networking technologies.

However, this new era brings its own set of challenges. The flexibility of these systems demands unmatched predictability and performance from the underlying network and timing infrastructures. The SMPTE ST 2110-21 standard[4] emphasizes this by setting strict timing and jitter requirements, explained even more clearly in the new RP 2110-25.[5] Factors like switch platforms' behavior particularities can influence the defined packet pacing benchmarks, highlighting the need for comprehensive testing before facility deployment.

While the Precision Time Protocol (PTP)[6] enables and ensures synchronization across all network and media devices, the allowed error budget for ST 2110 applications is rather tight. The overall timing performance of the system obviously relies on the individual performance and the introduced error of its individual components. In particular, the network components significantly impact the resulting quality of the time served to the media endpoints. Therefore, any lapse in network performance can jeopardize PTP's accuracy.

These considerations highlight the importance of test-

1545-0279/24©2024SMPTE

ing a switched platform's performance from the packet pacing and queueing, multicast replication, and timing performance under load before the full-scale deployment into an operational or commissioned facility. But despite the importance of such testing, it's often overlooked due to budget or time constraints and the availability of specialized equipment. Many broadcasters and integrators rely on manufacturers' claims, untested, inadequately tested, tested by rumor, tested with different platform designs, or tested using other software or firmware releases, risking unforeseen challenges.

To address this, the authors proposed an approach to media over IP network load testing,[7] emphasizing the use of accessible hardware and open-source software. This initiative aims to help broadcasters validate their networks, ensuring they meet the rigorous criteria imposed by relevant media standards and perform optimally.

> Creating a set of media profiles is essential in ensuring that testing closely mirrors real-world load, traffic patterns, and network scenarios, providing a common ground for testing and comparing system elements and ultimately guaranteeing more reliable and robust media over IP infrastructure.

Motivation for Further Research

While the authors proposed an accessible solution for media over IP network load testing using off-the-shelf hardware and open-source software, a significant gap outlined in the initial research remains. Though innovative in the methods but not new in nature, the approach surfaced the critical need for creating specific media traffic profiles.

The standard InternetMix (IMIX) profile,[8] commonly used in network testing, may not comprehensively represent media traffic's unique demands and characteristics. While IMIX offers a snapshot of general internet traffic, it lacks the granularity and specificity required to simulate the intricacies of media over IP traffic, such as the size of the flows, the load asymmetry due to the nature of media applications (e.g., studio vs. control room), the distribution of the "Elephant" and "Mice" streams,[9] the overwhelming presence of multicast traffic, etc. Thus, relying solely on traditional IMIX profiles could lead to an incomplete or inconclusive understanding of network performance under load, potentially overlooking critical bottlenecks or inefficiencies inherent to the nature of the media traffic.

Therefore, creating a set of media profiles is essential in ensuring that testing closely mirrors real-world load, traffic patterns, and network scenarios, providing a common ground for testing and comparing system elements and ultimately guaranteeing more reliable and robust media over IP infrastructure.

This observation provides the motivation for future research. There is a universally agreed need to investigate further the creation and optimization of the media-specific traffic profiles. Such profiles would provide more accurate representations of real-world media traffic patterns and offer insights into potential challenges and solutions specific to the media over IP domain. By building upon the foundational work previously done by the authors and emphasizing the creation of representative statistical media traffic profiles, this paper describes the research that lays the path forward for more reliable, efficient, and conclusive media-over-IP-oriented network testing, ensuring that the industry is well-equipped on its journey towards infrastructures supporting the next generation of media production applications.

Considerations for Effective ST 2110 Media Traffic Profile

As explained in the preceding sections, when load testing a network or a switch platform intended for use in the media applications involving the ST 2110 media streams, the efficacy of the test depends on the accuracy and completeness of the traffic profile in use. But what indeed makes a profile 'good'? Following multiple discussions with broadcasters, manufacturers, and system integrators during numerous industry forums, events, and conferences, the authors attempted to define the nuances of creating a representative ST 2110 media traffic profile, focusing on key characteristics that ensure precision in load testing.

A representative ST 2110 profile:

1545-0279/24©2024SMPTE

1. Distinguishes individual component sizes: Media traffic isn't homogenous. Due to their nature, different media essences or flows, video, audio, or metadata, will demonstrate different traffic patterns, even purely from a consumed bandwidth perspective. Recognizing and capturing these variations is crucial for building a meaningful statistical distribution model.
2. Accounts for bursts: While video essence is usually perceived as a steady, sustained load, a close inspection and fundamental understanding of the ST 2110-21 standard proves this perception devious.[10] Media traffic can be bursty, especially when the gapped profile is used. A good profile models these bursts accurately to simulate real-world conditions.
3. Puts things into perspective: Different production applications, scenarios, and types of produced content will demonstrate different distribution of media essence flows. A radio show will mostly show audio traffic. In contrast, a high-profile sports event will offer a mixture of live video camera feeds and synthetic senders like playout or graphics servers. The context of the profile is essential when applying it to a system designed for a particular application.
4. Considers aggregated impact: While differentiation of individual flows is necessary, the cumulative size of concurrent flows and its implications on network resources should be profiled. The profile should depict a ratio of an individual link or an overall network throughput versus the bandwidth consumed or even the ST 2110 traffic's share of the total network load in converged networks. This helps in understanding potential bottlenecks and prioritizing the media traffic where needed.
5. Assesses multicast behavior: Given ST 2110's reliance on multicast, its impact on network throughput, especially in terms of optimal layer 2 (IGMP)[11] and layer 3 (PIM)[12] performance, needs to be factored into the profile. While it is not trivial to replicate using synthetic traffic, the profile should strive to be representative, nonetheless.

An effective ST 2110 media traffic profile is more than just numbers. It represents the actual network behaviors in real-world conditions under various operational scenarios. It is unlikely that creating a "one size fits all" profile is possible. Instead, the authors concluded that a series of "one size fits most" profiles should be considered. By crafting a profile with careful attention to flow sizes, statistical distributions, and throughput considerations for various scenarios like radio shows, news shows, drama, sports or talk show productions, etc., network architects can ensure their load testing mirrors reality, leading to actionable insights and robust network designs.

To achieve this, the authors decided first to understand the options for collecting the required data from an operational system without impacting or impairing the network but also with minimal overhead for the operational staff. Therefore, a way for non-invasive traffic capturing in the critical points of the network needed to be identified, and these key points need to be identified. An easy-to-follow guide was required for the engineer performing the capturing exercise. The toolset for repeatable visualization and analysis of the data also needed to be defined. Lastly, the approach must be vendor-agnostic. The authors then decided to team up with broadcasters and system integrators of diverse scales worldwide to collect the required data. The authors' journey towards conducting the exercises is described in detail in the subsequent sections.

What Data is Required?

Authors' knowledge of typical production scenarios allows for a general understanding of the types of traffic generally present in the media networks.

This may include:

- ST 2110-20 uncompressed video streams of various formats (resolution, frame rate, chroma subsampling).
- ST 2110-22 compressed video flows of various formats (resolution, frame rate, chroma subsampling, codec type, compression rate).
- ST 2110-30 audio flows varying based on the number of channels, bit depth sampling rate, and packet time.
- ST 2110-31 audio flows, including PCM and non-PCM flows.
- ST 2110-40 ancillary data flows.
- PTP traffic.
- ST 2022-6 SDI-over-IP media flows.
- SRT or RIST flows.
- MPEG-TS flows.
- Other media traffic.
- Other non-media traffic.

Considering the criteria mentioned above for building a network traffic profile that would be as close to the real media production network as possible, one needs to determine which specific types of traffic from the list above are present.

For each of the traffic types, one also needs to know:

- The overall number of flows.
- The bandwidth of each flow and the total bandwidth consumed by this traffic type.
- The packet size and other flow characteristics, such as the protocol(s) used.

It is understood that despite commonalities, each specific media production network is unique. For example, using a different video format would significantly impact the bandwidth utilization in the network. The type of production will dictate the number of video flows versus the number of audio flows, thus dramatically changing the distribution of "elephant" and "mice" flows. As mentioned, finding a balance between generalizing the data to get closer to a universal network testing solution and keeping it specific enough for the cases that would be profoundly different from the others is not trivial. It is impossible to foresee every possible scenario, but the typical ones that significantly vary between each other can be identified.

For example, the following cases are compelling candidates for dedicated profiling:

- Media production facility with HD video and audio.

1545-0279/24©2024SMPTE

- Media production facility with UHD video and audio.
- Audio-only media production facility (e.g., a radio station).

Ways To Capture the Media Traffic Across a Live Network

Limitations of Port Mirroring for Non-Invasive Network Traffic Capturing

Port mirroring, often implemented as the Switched Port Analyzer (SPAN)[13] in many enterprise- and data center-grade switches, is a widely used method for capturing network traffic, especially when combined with a TAP aggregator.[14] While it offers advantages in specific scenarios, e.g., in corporate environments, where a real-time traffic analyzer can be used for early threat agent identifications, relying on port mirroring for non-invasive, low operational overhead network traffic capturing for high-bandwidth, media-centric standards like ST 2110, presents challenges.

Media networks often employ spine-leaf topology. A well-designed spine-leaf design carefully considers available links between spines and leaves. Port mirroring requires dedicating a switch port as the destination for mirrored traffic. This approach cannot be characterized as low overhead in environments where high-speed port availability is at a premium.

ST 2110 traffic, being media-centric, demands substantial bandwidth. Many capturing tools might not be equipped to handle the sheer volume of this traffic without data loss. While the authors previously suggested a solution for effective media traffic capturing,[7] it still requires equipment, time, and skills to set up.

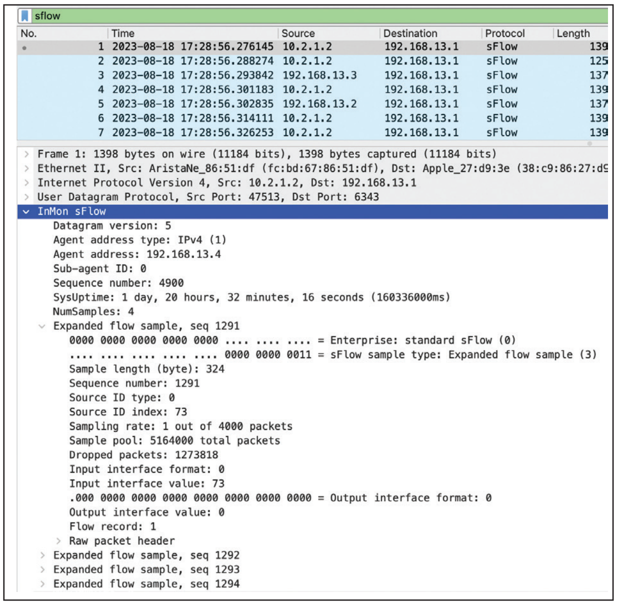
Finally, due to a combination of the previous two factors, depending on the scale of a system, capturing the entirety of ST 2110 traffic required to "snapshot" the system may require mirroring at multiple strategic points in the network. Not only does this increase the complexity of the capturing setup, but it also multiplies the challenges mentioned above, including port availability and potential network impairment.

A Viable Alternative

Considering the outlined challenges, the authors concluded that a lower-impact approach and toolset are required.

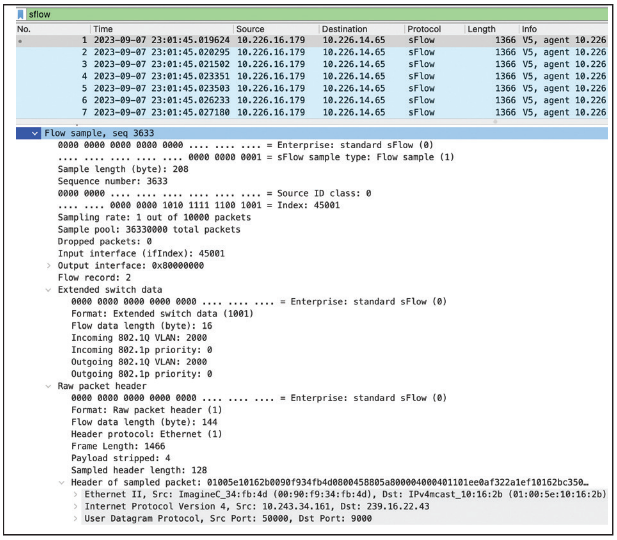
The appropriate method would gather only limited statistical data about the media network traffic instead of capturing the bulk of the traffic. The protocol selected for this task is sFlow,[15] described in more detail in the next section.

sFlow allows the switch to sample the traffic by capturing only one in a few thousand packets going through an interface. Furthermore, it won't capture the whole packet but rather a truncated version containing the headers and a few first bytes of the payload (depending on the implementation and configuration of the sFlow protocol). This allows for significantly lowering the utilization of resources needed to fulfill the task of gathering the data. Some switch platforms may have the sFlow protocol support implemented in hardware, effectively reducing the additional load generated by the required sampling. At the



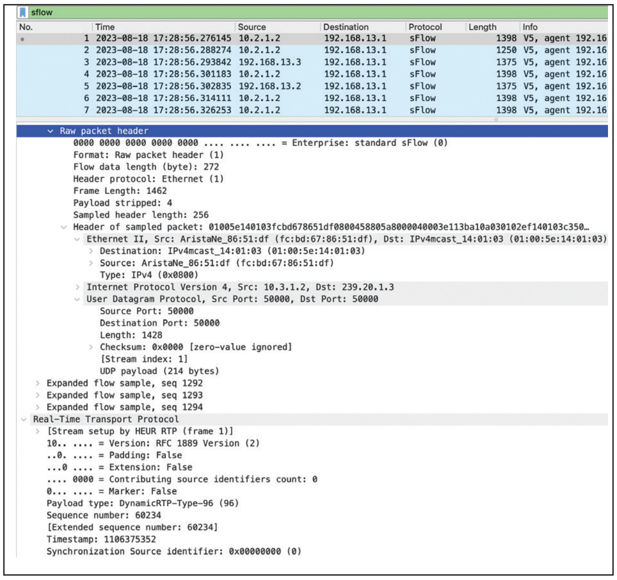
```
sflow
No.        Time                        Source         Destination    Protocol  Length
•        1 2023-08-18 17:28:56.276145  10.2.1.2       192.168.13.1   sFlow     139
         2 2023-08-18 17:28:56.288274  10.2.1.2       192.168.13.1   sFlow     125
         3 2023-08-18 17:28:56.293842  192.168.13.3   192.168.13.1   sFlow     137
         4 2023-08-18 17:28:56.301183  10.2.1.2       192.168.13.1   sFlow     139
         5 2023-08-18 17:28:56.302835  192.168.13.2   192.168.13.1   sFlow     137
         6 2023-08-18 17:28:56.314111  10.2.1.2       192.168.13.1   sFlow     139
         7 2023-08-18 17:28:56.326253  10.2.1.2       192.168.13.1   sFlow     139
> Frame 1: 1398 bytes on wire (11184 bits), 1398 bytes captured (11184 bits)
> Ethernet II, Src: AristaNe_86:51:df (fc:bd:67:86:51:df), Dst: Apple_27:d9:3e (38:c9:86:27:d9
> Internet Protocol Version 4, Src: 10.2.1.2, Dst: 192.168.13.1
> User Datagram Protocol, Src Port: 47513, Dst Port: 6343
⌄ InMon sFlow
    Datagram version: 5
    Agent address type: IPv4 (1)
    Agent address: 192.168.13.4
    Sub-agent ID: 0
    Sequence number: 4900
    SysUptime: 1 day, 20 hours, 32 minutes, 16 seconds (160336000ms)
    NumSamples: 4
  ⌄ Expanded flow sample, seq 1291
      0000 0000 0000 0000 0000 .... .... .... = Enterprise: standard sFlow (0)
      .... .... .... .... .... 0000 0000 0011 = sFlow sample type: Expanded flow sample (3)
      Sample length (byte): 324
      Sequence number: 1291
      Source ID type: 0
      Source ID index: 73
      Sampling rate: 1 out of 4000 packets
      Sample pool: 5164000 total packets
      Dropped packets: 1273818
      Input interface format: 0
      Input interface value: 73
      .000 0000 0000 0000 0000 0000 0000 0000 = Output interface format: 0
      Output interface value: 0
      Flow record: 1
    > Raw packet header
  > Expanded flow sample, seq 1292
  > Expanded flow sample, seq 1293
  > Expanded flow sample, seq 1294
```

FIGURE 1. sFlow basic data.



```
sflow
No.        Time                        Source         Destination    Protocol  Length  Info
•        1 2023-09-07 23:01:45.019624  10.226.16.179  10.226.14.65   sFlow     1366    V5, agent 10.226
         2 2023-09-07 23:01:45.020295  10.226.16.179  10.226.14.65   sFlow     1366    V5, agent 10.226
         3 2023-09-07 23:01:45.021502  10.226.16.179  10.226.14.65   sFlow     1366    V5, agent 10.226
         4 2023-09-07 23:01:45.023351  10.226.16.179  10.226.14.65   sFlow     1366    V5, agent 10.226
         5 2023-09-07 23:01:45.023503  10.226.16.179  10.226.14.65   sFlow     1366    V5, agent 10.226
         6 2023-09-07 23:01:45.026233  10.226.16.179  10.226.14.65   sFlow     1366    V5, agent 10.226
         7 2023-09-07 23:01:45.027180  10.226.16.179  10.226.14.65   sFlow     1366    V5, agent 10.226
  ⌄ Flow sample, seq 3633
      0000 0000 0000 0000 0000 .... .... .... = Enterprise: standard sFlow (0)
      .... .... .... .... .... 0000 0000 0001 = sFlow sample type: Flow sample (1)
      Sample length (byte): 208
      Sequence number: 3633
      0000 0000 .... .... .... .... .... .... = Source ID class: 0
      .... .... 0000 0000 1010 1111 1100 1001 = Index: 45001
      Sampling rate: 1 out of 10000 packets
      Sample pool: 36330000 total packets
      Dropped packets: 0
      Input interface (ifIndex): 45001
    > Output interface: 0x80000000
      Flow record: 2
    ⌄ Extended switch data
        0000 0000 0000 0000 0000 .... .... .... = Enterprise: standard sFlow (0)
        Format: Extended switch data (1001)
        Flow data length (byte): 16
        Incoming 802.1Q VLAN: 2000
        Incoming 802.1p priority: 0
        Outgoing 802.1Q VLAN: 2000
        Outgoing 802.1p priority: 0
    ⌄ Raw packet header
        0000 0000 0000 0000 0000 .... .... .... = Enterprise: standard sFlow (0)
        Format: Raw packet header (1)
        Flow data length (byte): 144
        Header protocol: Ethernet (1)
        Frame Length: 1466
        Payload stripped: 4
        Sampled header length: 128
      ⌄ Header of sampled packet: 01005e10162b0090f934fb4d0800458805a800004000401101ee0af322a1ef10162bc350…
        > Ethernet II, Src: ImagineC_34:fb:4d (00:90:f9:34:fb:4d), Dst: IPv4mcast_10:16:2b (01:00:5e:10:16:2b)
        > Internet Protocol Version 4, Src: 10.243.34.161, Dst: 239.16.22.43
        > User Datagram Protocol, Src Port: 50000, Dst Port: 9000
```

FIGURE 2. sFlow source and destination VLANs and priorities.



```
sflow
No.        Time                        Source         Destination    Protocol  Length  Info
•        1 2023-08-18 17:28:56.276145  10.2.1.2       192.168.13.1   sFlow     1398    V5, agent 192.16
         2 2023-08-18 17:28:56.288274  10.2.1.2       192.168.13.1   sFlow     1250    V5, agent 192.16
         3 2023-08-18 17:28:56.293842  192.168.13.3   192.168.13.1   sFlow     1375    V5, agent 192.16
         4 2023-08-18 17:28:56.301183  10.2.1.2       192.168.13.1   sFlow     1398    V5, agent 192.16
         5 2023-08-18 17:28:56.302835  192.168.13.2   192.168.13.1   sFlow     1375    V5, agent 192.16
         6 2023-08-18 17:28:56.314111  10.2.1.2       192.168.13.1   sFlow     1398    V5, agent 192.16
         7 2023-08-18 17:28:56.326253  10.2.1.2       192.168.13.1   sFlow     1398    V5, agent 192.16
    ⌄ Raw packet header
        0000 0000 0000 0000 0000 .... .... .... = Enterprise: standard sFlow (0)
        Format: Raw packet header (1)
        Flow data length (byte): 272
        Header protocol: Ethernet (1)
        Frame Length: 1462
        Payload stripped: 4
        Sampled header length: 256
      ⌄ Header of sampled packet: 01005e140103fcbd678651df0800458805a8000040003e113ba10a030102ef140103c350…
        ⌄ Ethernet II, Src: AristaNe_86:51:df (fc:bd:67:86:51:df), Dst: IPv4mcast_14:01:03 (01:00:5e:14:01:03)
          > Destination: IPv4mcast_14:01:03 (01:00:5e:14:01:03)
          > Source: AristaNe_86:51:df (fc:bd:67:86:51:df)
            Type: IPv4 (0x0800)
        > Internet Protocol Version 4, Src: 10.3.1.2, Dst: 239.20.1.3
        ⌄ User Datagram Protocol, Src Port: 50000, Dst Port: 50000
            Source Port: 50000
            Destination Port: 50000
            Length: 1428
          > Checksum: 0x0000 [zero-value ignored]
            [Stream index: 1]
            UDP payload (214 bytes)
  > Expanded flow sample, seq 1292
  > Expanded flow sample, seq 1293
  > Expanded flow sample, seq 1294
⌄ Real-Time Transport Protocol
  > [Stream setup by HEUR RTP (frame 1)]
    10.. .... = Version: RFC 1889 Version (2)
    ..0. .... = Padding: False
    ...0 .... = Extension: False
    .... 0000 = Contributing source identifiers count: 0
    0... .... = Marker: False
    Payload type: DynamicRTP-Type-96 (96)
    Sequence number: 60234
    [Extended sequence number: 60234]
    Timestamp: 1106375352
    Synchronization Source identifier: 0x00000000 (0)
```

FIGURE 3. Expanded flow sample, RAW packet header.

1545-0279/24©2024SMPTE

same time, with the extensive packet rates typically found in the media production network, even such sampled data should allow for building accurate traffic profiles.

The information in the sFlow sample is generally sufficient to place the sampled packet into one of the categories described above and help build accurate statistical traffic distribution profiles. As highlighted in previous work exploring the heuristic approaches used while developing the EBU LIST tool,[16] various combinations of audio and video formats in ST 2110 produce packets of specific sizes. Therefore, an assumption was made that knowing the packet size, overall bandwidth, RTP payload type, and overall description of formats used in the sampled network allows accurate categorizing of the flow. The exploration of this assumption is described in detail in the following sections. To verify or double-check the classification's accuracy, the authors plan to ask those providing the data about media formats expected to be present in their media production facilities.

The Proposed Toolset for Data Collection

sFlow Protocol

sFlow (sampled flow) is an industry-standard for exporting packet information at Layer 2. It was originally designed by InMon Corp and standardized via the IETF RFC 3176. The current version of sFlow is version 5.[17]

The main purpose of sFlow is to provide a means for exporting truncated packets alongside counters for interfaces.

To deliver at scale, sFlow requires the use of packet sampling. The most common method is random sampling (1 packet out of N). This is referred to as a "flow sample," while interface counter information is referred to as a "counter sample." Both are transmitted in the sFlow datagram from the host, switch, or router, capturing the samples to a centralized server known as the "sFlow collector."

As previously stated, flow samples are based on a defined sampling rate, i.e., an average of 1 out of N packets randomly sampled. The results are not 100% accurate but do provide a quantifiable accuracy, especially in the context of long-lived (aka, elephant) flows in SMPTE ST 2110-based systems.

sFlow datagrams are transmitted as port 6343 UDP packets to the specified host(s).

Within the sFlow packet, the following fields are contained:

- sFlow version
- Source IP address
- Sequence number
- Number of samples within the packet
- Flow and/or counter sample(s)

Additional switch/router vendor extensions may be included:

- Source and destination VLANs and priorities
- IP version and address of next-hop router
- Source and destination network mask lengths

Figures 1, 2, and 3 demonstrate how the sFlow data is interpreted in Wireshark.[18]

Additionally, vendor-specific hardware and/or software implementations allow for the sampling rate, max datagram size, and max header size (i.e., truncated sampled packet) to be selected. These may directly impact the specific sampling device performance based on the CPU, hardware offload, sampled interface packet rate, and other related implementation-specific performance parameters.

Hardware vs. Software sFlow Implementations

As mentioned above, the switch manufacturer may implement the sFlow protocol in hardware or software. The authors have evaluated a selection of platforms from three switch vendors most prominent in the media and entertainment segment. While the hardware implementation enjoys finer sampling granularity and offloads the required processing from a CPU to an ASIC, the initial testing shows that the software implementations produce acceptable results for this research.

Arista has several platforms that support sFlow in hardware. The rest of the Arista switch platforms support sFlow in software.

All NVIDIA/Mellanox switch platforms support sFlow in hardware.

Not all Cisco switch platforms support sFlow. Those that do (some of the Cisco Nexus 3000 and Cisco Nexus 9000 series) support it in software.

Depending on the implementation and the switch manufacturer's choices, sFlow, on some platforms, may only be able to sample ingress packets on an interface. In contrast, duplex traffic may be available for sampling on other platforms. This is an important consideration that needs to be considered when identifying the point in a network for meaningful data capturing.

The authors have conducted a lab experiment to collect sample reference data, test sFlow implementation on one of the switches from each vendor as a proof of concept, but also to stress test the switches to ensure that sFlow will not generate an excessive load of the switch resources, considering that the data will be gathered predominantly in the production networks. This experiment is described and analyzed in the following sections.

sFlow Collector

Typically, in traditional enterprise IT or data center environments where means for detecting suspicious traffic or traffic behavior is needed, an sFlow collector is used. sFlow collector is a tool that aggregates sFlow data from multiple agents (samplers)[19] in the network. It can be a dedicated proprietary or open-source software tool or embedded into a larger control and monitoring solution.

Depending on the implemented feature set, an sFlow collector can be advanced, aware of the network topology, and attempt to provide meaningful analytics and intelligence. Alternatively, it can simply do the initial processing of the available sFlow sampled data and expose it to another analytics engine.

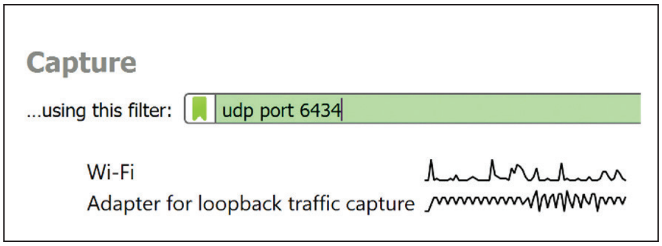
During the preparatory phase of this research, the authors have tested numerous sFlow collectors in an attempt

1545-0279/24©2024SMPTE

to find one that is able to parse the sFlow packets coming from the agents simply, extract as much collected data as available in the packets, and expose it in as a raw form as possible. Unfortunately, the optimal tool, neither free nor commercial, open, or closed source, wasn't identified, and an alternative approach was required.

Alternative sFlow Data Collection

The authors suggest capturing the raw sFlow traffic on any available host in the network accessible from the switch(es) where sFlow will be configured. It is important to remember that most switch platforms output sFlow packets via a selected front panel interface rather than a management interface. The capturing host must be able to capture the traffic without packet loss due to kernel or other pipeline limitations. A Linux host is preferable over a Windows host. Tcpdump,[20] Wireshark, or any other network sniffer that can capture in PCAP format will work. The authors recommend looking into one of their previous works[21] that describes a high-performance packet capturing method using off-the-shelf equipment. When using Wireshark, the authors suggest using a packet capture filter (not to be confused with a display filter) so that only sFlow traffic is captured. An example is shown in **Fig. 4**.



Capture

...using this filter: udp port 6434

Wi-Fi

Adapter for loopback traffic capture

FIGURE 4. Example of a Wireshark packet capture filter.

The toolset explained in the following sections expects a standard uncompressed capture,[22] so the captured data must be saved in an uncompressed PCAP format rather than PCAPNG. Due to the "one in n packets" sampling nature of the sFlow, the authors recommend performing a long-duration capture (e.g., a whole one-hour period) so that the streams with fewer packets per second will be sufficiently sampled.

Data Capturing Workflow

The overall high-level workflow for effective, non-invasive, and meaningful data capturing includes:

- Selecting the switch(es) where the data will be collected.
- Determining the link(s) on this(-ese) switch(es) to get the data from.

Depending on the sFlow implementation (hardware vs. software, ingress only vs. duplex), selecting points to sample is imperative to the quality of data sampling and its ability to provide a representative overview of the traffic patterns in a network.

A sample simplified diagram of a typical spine-leaf topology-based media network is shown in **Fig. 5**.

Depending on the applications and production scenarios, but also on operational limitations, assumptions and compromises on where to capture the traffic samples must be made. Ideally, all ports on both sides of all inter-switch links should be sampled. Due to the ability of the analysis toolset, described in the following section, to aggregate sFlow samples from multiple points and correlate the flows based on the topology, this approach gives the best visibility of the traffic passing through the networks.

Alternatively, traffic can be sampled based on the application. For example, traffic between a studio and a gallery/control room may give an informative representation of stream distribution for a particular production scenario. Considering the possible limitation of ingress-only sampling, it is advisable to sample on both sides of the inter-switch links of the involved leaves. One also needs to be mindful that in the multi-spine topologies with ECMP,[23] the traffic from one leaf may be spread across multiple spines. Unless SDN is being used, there may not be a reliable way to predict such scenarios. Therefore, sampling of all spines is advised.

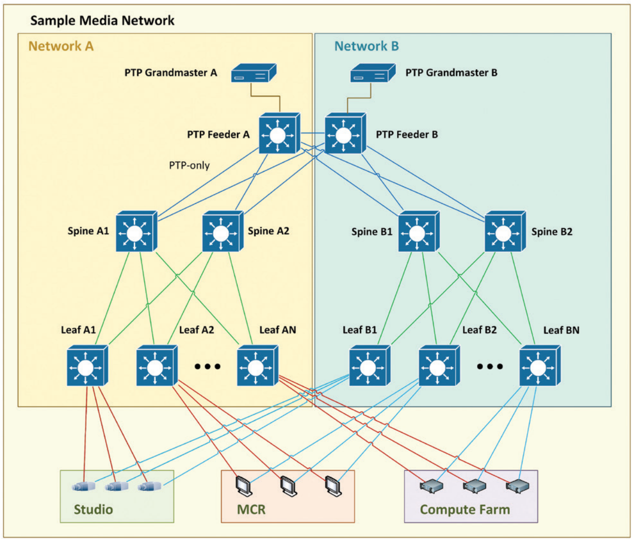


FIGURE 5. Simplified diagram of a spine-leaf-based media network.

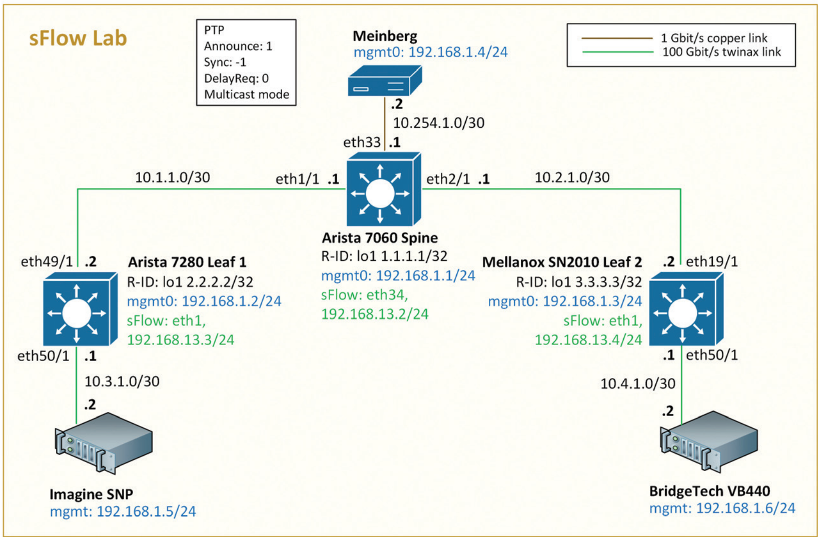


FIGURE 6. The EBU Lab sFlow testing setup.

- Configuring sFlow on the switch(es) on selected interfaces. The authors have assembled a basic set of commands for the three switch manufacturers listed above to enable sFlow on the identified interfaces. The snippets are available in the SMPTE 2023 Media Technology Summit (MTS) proceedings.
- Initiate a packet capture of sFlow traffic on a host in the network. As explained in the section "Alternative sFlow data collection," using a dedicated sFlow collector is not advised, and a simple packet capture suffices.

From a practical standpoint, the authors also suggest to:

- Perform the sampling measurements on an actual operational production network, preferably at peak usage time during typical operations.
- Use the highest possible sampling rate, producing the maximum number of traffic samples, depending on the platform, that may be one packet in 4,000 or one packet in 16,384.
- Provide additional information describing the general topology and setup of a facility without exact details about the vendor(s) and model(s) of the devices used, but preferably providing information on the video and audio formats used in the media production facility. The authors have compiled a checklist allowing the team to communicate this information with minimal admin overhead. It is provided in the SMPTE 2023 MTS proceedings.
- Capture as much sampled data as possible for as long as possible. Although even the data covering only a few minutes of samples will provide an initial set of information to better understand the overall picture in a network, as discussed in the previous sections, due to the sampling nature of the sFlow, the precision of capturing the less prominent streams like ST 2110-30 and ST 2110-40 benefits from the duration of sampling process (i.e., more packets of smaller streams are captured).

TABLE 1. Calculated bandwidth per video format.

#	Frame Width	Frame Height	Frame Rate (Hz)	Bit-Depth	Chroma Subsampling	Bandwidth (Gbits/s)
1	1920	1080	25	10	4:2:2	1,11
2	1920	1080	50	10	4:2:2	2,21
3	1920	1080	60	10	4:2:2	2,65
4	3840	2160	50	10	4:2:2	8,86
5	3840	2160	60	10	4:2:2	10,62

TABLE 2. Calculated bandwidth per audio format.

#	Sampling Rate (Hz)	Bit	Packet Time	Channel count	Payload (Bytes)	Bandwidth (Mbits/s)
1	48000	24	1ms	1	144	1,51
2	48000	24	1ms	2	288	2,61
3	48000	24	1ms	4	567	4,81
4	48000	24	1ms	6	864	7,00
5	48000	24	1ms	8	1152	9,20
6	48000	24	125µs	1	18	4,39
7	48000	24	125µs	2	36	5,49
8	48000	24	125µs	4	72	7,69
9	48000	24	125µs	6	108	9,89
10	48000	24	125µs	8	144	12,08
11	48000	24	125µs	16	288	20,87
12	48000	24	125µs	32	576	38,45
13	48000	24	125µs	64	1152	73,61

Analyzing the Data

EBU Lab sFlow Testing Experiment

To establish the baseline "fingerprint" of typical media flows and validate the fundamental assumption that a media flow can be identified from the data provided by the sFlow sampling process, the authors have conducted a series of lab experiments using a rudimentary media network and a set of endpoints consisting of:

- Arista 7060CX-32S spine switch[24]
- Arista 7280 leaf switch[25]
- NVIDIA SN2010 leaf switch[26]
- Imagine Communications Selenio Network Processor (SNP)[27]
- Bridge Technologies VB440 Network Analyzer[28]
- Meinberg Lantime M1000[29]

The diagram of the lab setup is illustrated in **Fig. 6**.

A series of reference video (ST 2110-20[30] and ST 2110-22[31] with JPEG XS[32]), audio (ST 2110-30[33] and ST 2110-31[34]) and ancillary (ST 2110-40[35]) streams were configured on the SNP, acting as a sender and received by the VB440, thus traversing the network from Leaf 1, via the Spine to Leaf 2. A list of reference flows configured on the SNP to establish a baseline is provided in the SMPTE 2023 MTS proceedings. sFlow sampling was configured in the key points in the

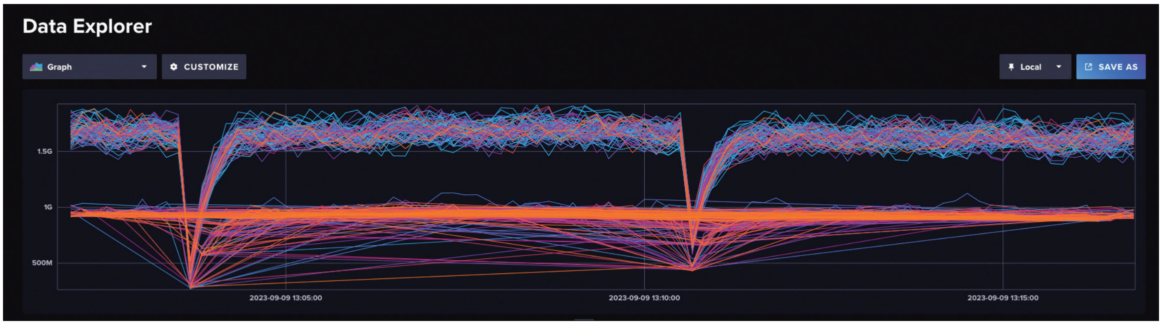


FIGURE 7. Influx Data Explorer.

1545-0279/24©2024SMPTE

network. The platforms only supported ingress sampling, with Arista Spine and Leaf 1 supporting the sampling in software and NVIDIA/Mellanox supporting it in hardware. Switch and router extensions were enabled where available (only on Arista switches), while the BGP extensions were turned off. A line port for sending sFlow data was allocated on every switch, and a PCAP capture was performed with a standard laptop (Apple MacBook Pro).

Expected Bandwidth Per Flow

The expected bandwidth requirements for the video streams must encompass the authentic bandwidth essential for transmitting the video data, containing the Ethernet frame overhead as shown in **Table 1** (calculated bandwidth per video format). Notably, the GAP necessary for adhering to the stringent requirements of the ST 2110-21 profile cannot be measured with the identified toolset (explained in detail further) yet.

The expected bandwidth for audio streams should include both the PCM audio data and the Ethernet frame overhead, as **Table 2** details the calculated bandwidth for a limited amount of audio formats. It is essential to recognize that this overhead component can indeed be quite substantial in the case of audio. The audio bandwidth is calculated, including the overhead needed for the Ethernet frame.

Interpreting Raw sFlow Data

It appears that a programmatic approach is needed to access the raw data from the sFlow packets without employing pre-defined analytics. Additionally, since sFlow is sampling the packets over time, a solution could extract the data into a time series database.

InfluxDB[36] was identified as a suitable candidate for a time series database that can run custom queries against time series datasets like streaming telemetry. Starting from version 2, InfluxDB includes a feature-rich data visualization toolset, a dedicated scripting language, and the option to export time sections as CSV files. However, the essential feature for this research was the introduced native support for some types of streaming telemetry scrapping, particularly in Prometheus[37] format, as well as an extensive list of input data transformers. InfluxDB can be used as a cloud-based SaaS service. However, for the data volumes involved in the study, the authors have successfully used the Docker-based InfluxDB container.

InMon, the company behind the initial definition of sFlow, maintains a reference implementation of an sFlow collector, the sFlow-RT.[38] While the native data analytics are limited, sFlow-RT has two crucial features that became groundbreaking for this research.

sFlow-RT can replay the PCAP files containing sFlow packets and present the data to external consumers. Additionally, sFlow-RT has native support for translating the exposed data into Prometheus format, thus making it directly accessible to the standard InfluxDB scrapper without requiring input processing.[39,40]

Thanks to extensive documentation[41] and some tinkering with the API definitions, the authors have assembled a toolset based on the sFlow-RT and InfluxDB contains that can be deployed via Docker-compose.[42] The deployment guide is available in the SMPTE MTS 2023 proceeding.

After pointing the sFlow-RT to a PCAP to be analyzed and launching both containers, the PCAP replay is started automatically, and the data is pulled into the defined InfluxDB buckets for further processing.

Analyzing the LAB Streams

After completing the toolset preparation and setting up an experimental environment, a reference sFlow PCAP with sampled streams was created for 10 min. The data pattern resulting from the replay looping feature of sFlow-RT is clearly visible in the Influx Data Explorer shown in **Fig. 7**. As mentioned, the goal of this part of the exercise is to validate the assumption that the media stream type can be identified, as well as to attempt to identify, cross-check, and "fingerprint" as many streams as possible from the reference capture, to be able to recognize the traffic patterns of different types of streams in the production captures. This allows for a faster and more scalable definition of the statistical models.

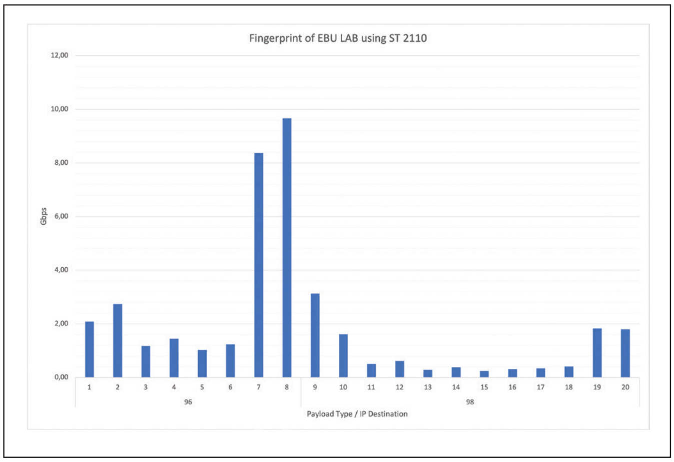


FIGURE 8. ST 2110 video streams collected by sFlow at the EBU Lab.

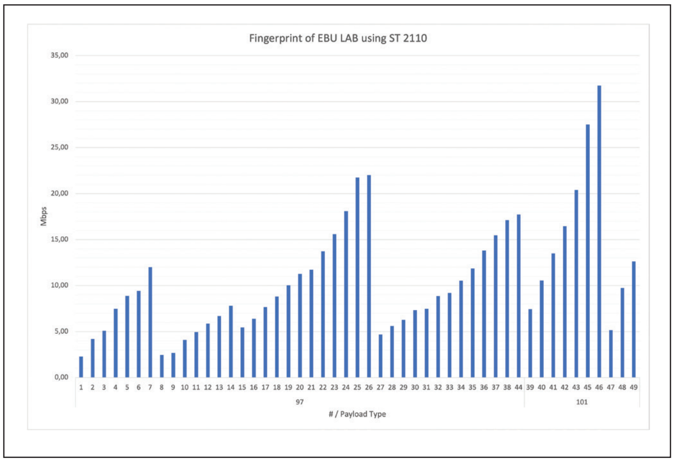


FIGURE 9. ST 2110 audio streams collected by sFlow at the EBU Lab.

TABLE 3. Video streams provided in the EBU Lab PCAP file.

#	Payload Type	IP Destination	Reported Bandwidth β_1 (Gbps)	Picture Height (px)	Scan	Frame Rate (Hz)	Standard	Measured Bandwidth β_2 (Gbps)	β_1/β_2
1	96	239.20.1.1	2,09	1080	P	50	ST 2110-20	2,17	-4%
2	96	239.20.1.2	2,73	1080	P	60	ST 2110-20	2,60	5%
3	96	239.20.2.1	1,18	1080	I	50	ST 2110-20	1,08	9%
4	96	239.20.2.2	1,45	1080	I	60	ST 2110-20	1,30	12%
5	96	239.20.3.1	1,03	720	P	50	ST 2110-20	0,96	7%
6	96	239.20.3.2	1,24	720	P	60	ST 2110-20	1,16	7%
7	96	239.20.5.1	8,60	2160	P	50	ST 2110-20	8,67	-1%
8	96	239.20.5.2	9,66	2160	P	60	ST 2110-20	10,40	-7%
9	98	239.200.1.1	3,13	1080	P	50	ST 2022-6	3,10	1%
10	98	239.200.1.2	1,62	1080	I	50	ST 2022-6	1,55	5%
11	98	239.22.1.1	0,50	1080	P	50	ST 2110-22	0,45	11%
12	98	239.22.1.2	0,63	1080	P	60	ST 2110-22	0,54	17%
13	98	239.22.2.1	0,29	1080	I	50	ST 2110-22	0,23	26%
14	98	239.22.2.2	0,38	1080	I	60	ST 2110-22	0,27	41%
15	98	239.22.3.1	0,26	1080	P	50	ST 2110-22	0,21	24%
16	98	239.22.3.2	0,35	720	P	60	ST 2110-22	0,24	46%
17	98	239.22.4.1	0,34	1080	P	50	ST 2110-22	0,27	26%
18	98	239.22.4.2	0,41	1080	P	60	ST 2110-22	0,32	28%
19	98	239.22.5.1	1,83	2160	P	50	ST 2110-22	1,81	1%
20	98	239.22.5.2	1,80	2160	P	60	ST 2110-22	1,81	-1%

The sFlow-RT parsed data is divided into three independent buckets based on the way sFlow presents the data internally: “sflow-all-flows,” “sflow-analyzer,” and “sflow-metrics.” The first bucket, “sflow-all-flows,” is particularly important since it contains detailed information on the sampled flows. The authors chose to export this data in CSV format due to available experience with Microsoft Excel’s analytical tools. When importing CSV data, validating its representation and guaranteeing appropriate translation into the needed Excel format is expected.

Video Streams

In the lab setup, a total of 20 distinct video streams were established using an Imagine SNP, each with a unique format. The objective behind this setup was to generate predetermined flows, enabling a clear distinction between the manually generated data and what the tooling could autonomously identify. On top of the sFlow collector, all the generated streams were measured using a well-established tool used in the industry, a Bridge Technologies VB440. **Figure 8** illustrates that all the produced streams were captured, stored in the PCAP file, run through the tool, and visualized with Excel. Access to key parameters such as Payload Type, IP destination address, and bandwidth allow for deeper analysis. The decision was made to display the different multicast addresses ordered per Payload Type with the bandwidth consumed on the network.

As commonly understood, media streams lack inherent self-description. Without the Session Description Protocol (SDP)[43] or any prior knowledge of network activities, distinguishing between different streams can be a formidable task. At first glance, recognizing distinct video formats might seem uncomplicated. However, grasping the essence of the media flow can be daunting without prior familiarity. Notably, certain critical fields, such as the RTP timestamp, sequence number, and Marker Bit, remain inaccessible with the current tooling.

Table 3 shows the 20 different video streams with the payload type (PT), multicast address (IP DST), bandwidth reported using sFlow (β_1), horizontal resolution (H), scanning (progressive or interlaced), frame rate (FR), which standard was used, bandwidth reported by the VB440 (β_2), and last value represents the deviation between β_1 and β_2 in percentage.

Audio Streams

In addition to the 20 video streams, 49 unique audio streams were generated. **Figure 9** illustrates all the bandwidth consumed per multicast stream, ordered per Payload Type.

Table 4 shows these streams each with their payload type (PT), multicast address (IP DST), bandwidth reported using sFlow (β_1), Sample Rate (SR), Bit-Depth (Bit), Packet Time (P.Time), amount of audio channels per stream (CH), bandwidth reported by the VB440 (β_2), and a value representing the deviation between β_1 and β_2 as a percentage.

The comparison between β_1 and β_2 demonstrates the limitation of sFlow in the accuracy of capturing the bandwidth. Based on the statistical nature of sFlow and the low bandwidth of audio streams, the current hypothesis is that a longer capture time is required to create a more accurate result. This will be subject to further testing.

Ancillary Streams

Taking a closer look at the measurements of the ST 2110-40 streams provided in the EBU lab, significant disparities become evident. However, it’s important to note that all the provided ST 2110-40 streams were identical. The maximum value, as illustrated in **Fig. 10**, is twice the magnitude of the minimum value. This stark contrast vividly

TABLE 4. Audio streams provided in the EBU Lab PCAP file.

#	Payload Type	IP Destination	Reported Bandwidth β_1 (Mbps)	Sampling Rate (Hz)	Bit depth	Packet Time	Channel count	Measured Bandwidth β_2 (Mbps)	β_1/β_2
1	97	239.30.100.1	2,41	48000	24	1ms	1	1,58	53%
2	97	239.30.100.2	3,89	48000	24	1ms	2	2,74	42%
3	97	239.30.100.3	5,47	48000	24	1ms	3	3,89	41%
4	97	239.30.100.4	6,98	48000	24	1ms	4	5,04	38%
5	97	239.30.100.5	8,83	48000	24	1ms	5	6,19	43%
6	97	239.30.100.6	8,99	48000	24	1ms	6	7,34	22%
7	97	239.30.100.7	12,60	48000	24	1ms	7	8,50	48%
8	97	239.30.101.1	2,43	48000	16	1ms	1	1,97	23%
9	97	239.30.101.2	2,78	48000	16	1ms	2	1,97	41%
10	97	239.30.101.3	4,11	48000	16	1ms	3	2,74	50%
11	97	239.30.101.4	4,63	48000	16	1ms	4	3,50	32%
12	97	239.30.101.5	5,67	48000	16	1ms	5	4,27	33%
13	97	239.30.101.6	6,53	48000	16	1ms	6	5,04	30%
14	97	239.30.101.7	7,34	48000	16	1ms	7	5,81	26%
15	97	239.30.125.1	5,27	48000	24	125µs	1	4,61	14%
16	97	239.30.125.2	6,27	48000	24	125µs	2	5,76	9%
17	97	239.30.125.3	7,51	48000	24	125µs	3	6,92	9%
18	97	239.30.125.4	8,72	48000	24	125µs	4	8,06	8%
19	97	239.30.125.5	9,87	48000	24	125µs	5	9,22	7%
20	97	239.30.125.6	11,07	48000	24	125µs	6	10,37	7%
21	97	239.30.125.7	11,55	48000	24	125µs	7	11,52	0%
22	97	239.30.125.8	13,49	48000	24	125µs	8	12,67	6%
23	97	239.30.125.9	15,25	48000	24	125µs	10	14,98	2%
24	97	239.30.125.10	17,99	48000	24	125µs	12	17,28	4%
25	97	239.30.125.11	21,20	48000	24	125µs	14	19,58	8%
26	97	239.30.125.12	21,55	48000	24	125µs	16	21,89	-2%
27	97	239.30.126.1	4,51	48000	16	125µs	1	4,22	7%
28	97	239.30.126.2	5,35	48000	16	125µs	2	4,99	7%
29	97	239.30.126.3	6,16	48000	16	125µs	3	5,76	7%
30	97	239.30.126.4	7,16	48000	16	125µs	4	6,53	10%
31	97	239.30.126.5	7,28	48000	16	125µs	5	7,30	0%
32	97	239.30.126.6	8,62	48000	16	125µs	6	8,06	7%
33	97	239.30.126.7	9,12	48000	16	125µs	7	8,83	3%
34	97	239.30.126.8	10,02	48000	16	125µs	8	9,60	4%
35	97	239.30.126.9	11,38	48000	16	125µs	10	11,14	2%
36	97	239.30.126.10	13,55	48000	16	125µs	12	12,67	7%
37	97	239.30.126.11	15,13	48000	16	125µs	14	14,21	6%
38	97	239.30.126.12	16,49	48000	16	125µs	16	15,74	5%
39	101	239.31.125.1	7,19	48000	32	125µs	2	6,53	10%
40	101	239.31.125.2	10,23	48000	32	125µs	4	9,60	7%
41	101	239.31.125.3	13,20	48000	32	125µs	6	12,67	4%
42	101	239.31.125.4	16,04	48000	32	125µs	8	15,74	2%
43	101	239.31.125.5	19,89	48000	32	125µs	10	18,82	6%
44	97	239.31.125.6	17,02	48000	24	125µs	12	17,28	-2%
45	101	239.31.125.7	26,31	48000	32	125µs	14	24,96	5%
46	101	239.31.125.8	30,83	48000	32	125µs	16	28,03	10%
47	101	239.31.100.1	5,35	48000	32	1ms	2	3,50	53%
48	101	239.31.100.2	9,45	48000	32	1ms	4	6,58	44%
49	101	239.31.100.3	12,79	48000	32	1ms	6	9,65	33%

underscores the constraints of conducting relatively brief captures using a sampling method like sFlow. Further testing is required to establish a more accurate image of the "mice" flows present in the network.

Combined Streams

All lab-generated streams consume over 36 Gbits/s on average. This data throughput underscores the intensive nature of the streaming activities being analyzed. **Table 5** provides a breakdown of the various media streams per payload type (PT), shedding light on each stream's bandwidth (**β**) distribution within the experiment.

It's worth noting that in this experimental setup, all streams have been deliberately configured to utilize dis-

1545-0279/24©2024SMPTE

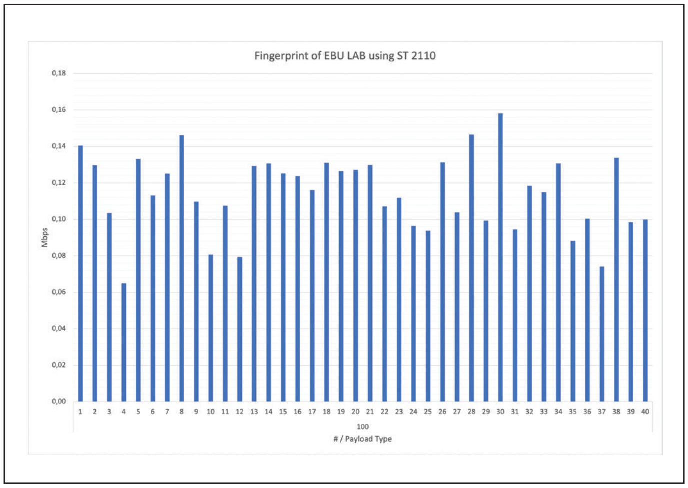


FIGURE 10. ST 2110 ancillary streams collected by sFlow at the EBU Lab.

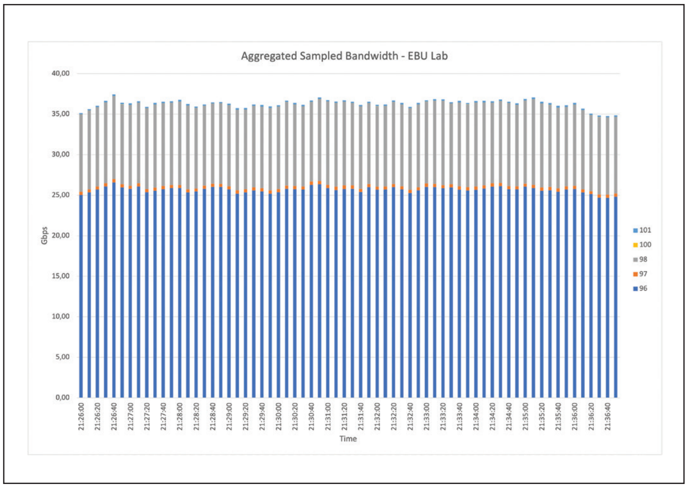


FIGURE 11. Aggregated sampled bandwidth collected by sFlow at the EBU Lab.

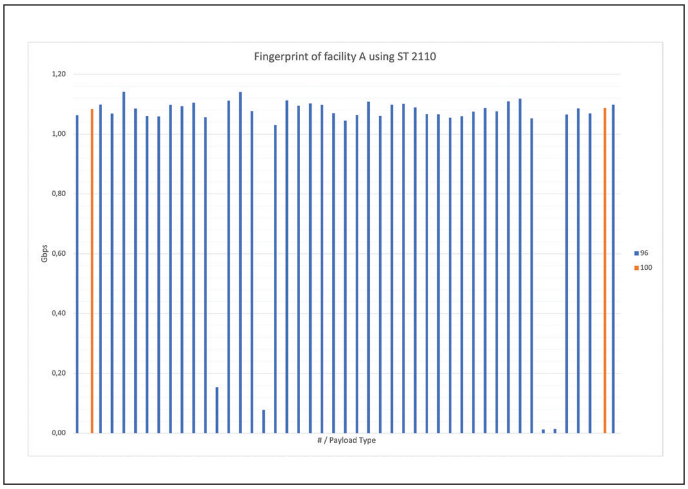


FIGURE 12. ST 2110 video streams collected by sFlow at Facility A.

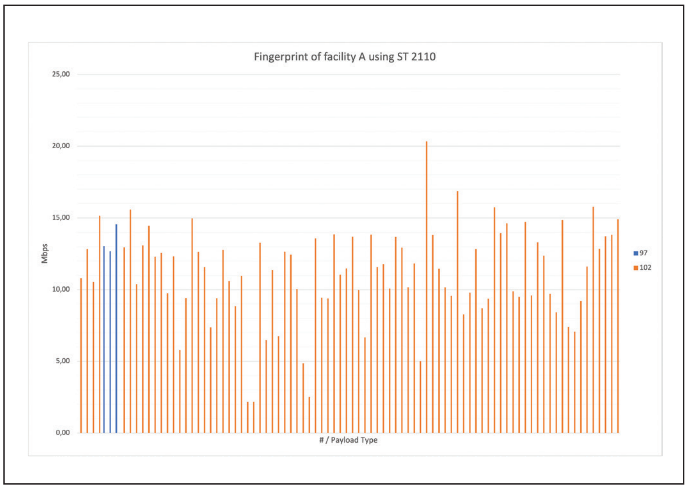


FIGURE 13. ST 2110 audio streams collected by sFlow at Facility A.

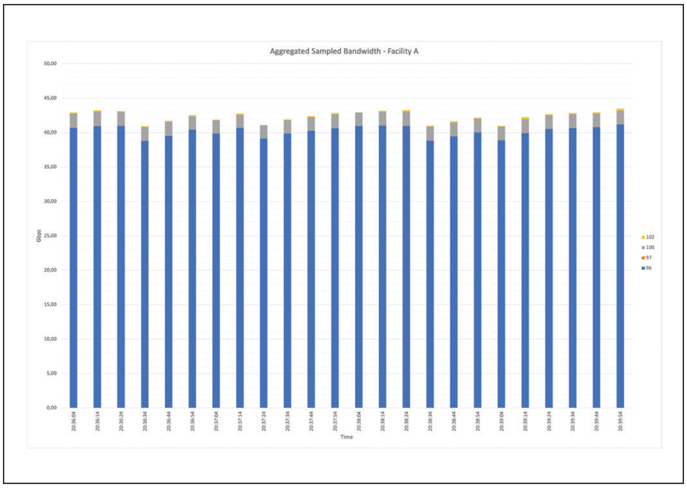


FIGURE 14. Aggregated sampled bandwidth collected by sFlow at Facility A.

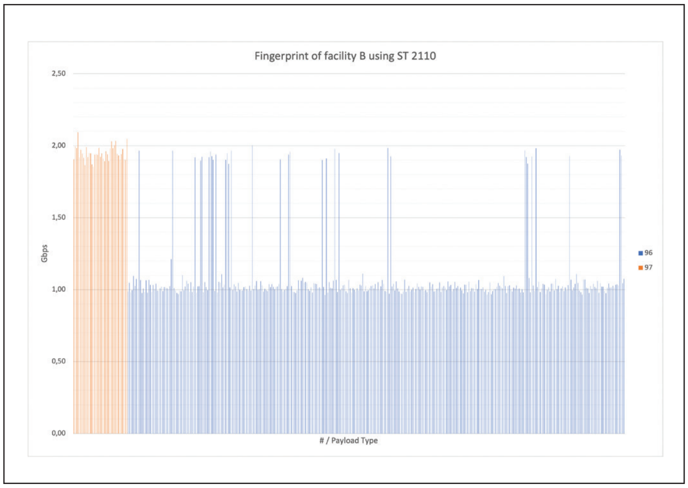


FIGURE 15. Fingerprint of facility B.

tinct formats. The decision was made to stress-test the system and assess its performance under challenging conditions. However, it's essential to acknowledge that such a scenario, where every stream has a different format, differs from a typical representation of real-world facilities. In practical environments, there is usually a greater degree of standardization in media formats to ensure smoother operations and interoperability.

The duration of the captured data was 10 min. Ideally, one would anticipate a consistent and stable data flow pat-

1545-0279/24©2024SMPTE

tern within this fixed timeframe. However, there is a noticeable variation over time within this window, as shown in **Fig. 11**, which is unexpected.

One possible explanation for this variation could be attributed to the sampling method employed in collecting sFlow data. It may necessitate a closer examination of the sFlow sampling methodology and a more thorough investigation into the factors causing the observed deviations in the data transmission pattern.

Analyzing the PCAP of a Real Facility - A

After scrutinizing the configuration and thoroughly assessing the EBU Lab data, the authors started a comprehensive evaluation of the first real-world PCAP file that had been graciously provided by one of the broadcasters. Accompanying this PCAP file was a brief note, which briefly stated the expected traffic characteristics: 32 Gbits/s of traffic should be directed toward the spine, while 15 Gbits/s should be flowing from the spine in this specific capture. Furthermore, it was emphasized that all flows within this dataset were uncompressed 1080i/25 streams, adding a critical contextual dimension to our analysis.

Employing the same methodology outlined earlier, the data was extracted from the PCAP file and transformed into an Excel format to facilitate further in-depth analysis. The sampled data contains a duration of three minutes within the capture.

Video Streams

Figure 12 demonstrates the traffic encapsulated in the PCAP file with payload types 96 and 100.

Without further knowledge of the facility, one can assume this is all ST 2110-20 traffic in the mentioned 1080i/25 formats. It seems that 4 flows are being terminated or initiated during the capture.

Audio Streams

Figure 13 illustrates the traffic contained within the PCAP file, explicitly focusing on payload types 97 and 102. There are 85 flows characterized by payload type 102 within this dataset, identified as ST 2110-30 audio streams. These streams consist of eight audio channels with a packing time of 125µs, resulting in a data rate of 12 Mbits/s.

Ancillary Streams

No ancillary streams were found in this capture.

Combined Streams

The distribution of the different flows is summarized in **Table 6** and visualized in **Fig. 14**.

Analyzing the PCAP of a Real Facility - B

At the time of writing, the authors have received another sample PCAP file, generously shared by a different broadcaster. This capture appears to contain 522 media streams distributed across three distinct payload types. The majority of these streams are configured as 1080i/25 video

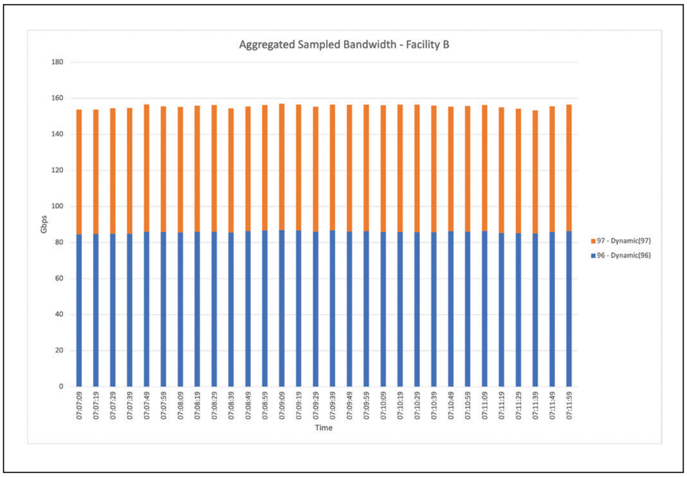


FIGURE 16. Aggregated Sampled Bandwidth collected by sFlow at Facility B.

Glossary of Acronyms	
AES67	Audio Engineering Society Standard 67
API	Application Programming Interface
ASIC	Application-Specific Integrated Circuit
BGP	Border Gateway Protocol
CPU	Central Processing Unit
CSV	Coma Separated Value
EBU LIST	European Broadcasting Union’s Live Ip Software Toolkit
ECMP	Equal-Cost Multi-Path (Routing)
HD	High Definition
IEEE	Institute of Electrical and Electronics Engineers
IETF	Internet Engineering Task Force
IGMP	Internet Group Management Protocol
IMIX	Internet Mix
IP	Internet Protocol
IT	Information Technology
MPEG-TS	Moving Picture Experts Group’s Transport Stream
PCM	Pulse-Code Modulation
PIM	Protocol Independent Multicast
PTP	Precision Time Protocol
RFC	Request For Comments
RIST	Reliable Internet Stream Transport
RP	Recommended Practice
RTP	Real-Time Protocol
SAAS	Software As a Service
SDI	Serial Digital Interface
SDN	Software Defined Network
SPAN	Switched Port Analyzer
SRT	Secure Reliable Transport
ST	Standard
TAP	Test Access Point
UHD	Ultra-High Definition
VLAN	Virtual Local Area Network

TABLE 5. Overview of traffic distribution.

#	Payload Type	Media type (SMPTE Standard)	Streams	Bandwidth (Gbps)	Distribution
1	96	ST 2110-20	8	25,682	70,74%
2	97	ST 2110-30	39	0,392	1,08%
3	98	ST 2110-22 / ST 2022-6	12	10,044	27,67%
4	100	ST 2110-40	40	0,002	0,01%
5	101	ST 2110-31	10	0,182	0,50%

TABLE 6. Overview of traffic distribution

#	Payload Type	Media type (SMPTE Standard)	Streams	Bandwidth (Gbps)	Distribution
1	96	ST 2110-20	45	44,68	93,40%
2	97	ST 2110-30	3	0,04	0,08%
3	100	ST 2110-20	2	2,17	4,54%
4	102	ST 2110-30	85	0,95	1,98%

streams. It's worth noting that a complete snapshot of the facility was not provided, as no audio streams were detected during our initial analysis. Following the same established methodology, the data was extracted from the PCAP file and converted into an Excel format to conduct a more comprehensive and detailed analysis.

The sampled data in this capture spans over four minutes and is visualized in **Fig. 15**.

Combined Streams

Unfortunately, the fact that only the video essence streams were sampled in this capture renders the results non-representative, as 100% of the used bandwidth is consumed by the elephant flows, as demonstrated in **Fig. 16**. This proves that a complete snapshot of all media traffic in the facility is required to be able to extract the distribution model, no matter how numerous those flows are. Therefore, the capturing exercise must be repeated in facility B.

Conclusion and Future Work

One of the primary goals of this research was to identify a method for non-intrusive and low-overhead capturing of high volumes of the ST 2110 network traffic and to develop a methodology for its analysis. It is also essential to highlight the challenges faced during this research. The complexities of media traffic and the nuances in its patterns made the analysis a daunting task, taking longer than the authors anticipated.

While the proposed comprehensive toolset offers a robust approach to capturing and analyzing network traffic, the challenge lies in interpreting this vast amount of data to draw meaningful conclusions once the PCAP is obtained. The data provides a selection of available parameters, from packet sizes to RTP payload types, which can be used for categorizing flows and understanding the nature of the traffic. However, as ST 2110 essence streams use dynamic payload types and the actual values are mere recommendations, a stream may be misidentified should an implementation use another payload type. This raises concerns about the potential need to process the sFlow packets[44,45] directly using a custom processing pipeline. This could allow for extracting additional valuable information, such as RTP headers.

The achieved results validate some initial assumptions, and the insights gained are invaluable in taking next steps toward developing a comprehensive media IMIX profile for the media industry. However, the question remains: Do longer captures provide a more granular and accurate bandwidth measurement? Initial findings suggest this might be the case, but further validation is required. Notably, all packets in various essence types seem uniformly distributed, except for the last packet of a video stream (-20 and -22), which is smaller. These new assumptions need further exploration as more data is gathered.

Future work will focus on refining the analysis methodology, validating the findings, and compiling and fine-tuning the IMIX profiles. At the time of writing, the authors actively seek additional participants willing to provide traffic snapshots of their facilities. Broader and more diverse data will allow for more representative profiles. The readers are encouraged to contact the authors via 2110-media-profile@list.ebu.ch.

Acknowledgments

These projects would never have been possible without the excellent collaboration between the European Broadcasting Union, NVIDIA, and the Vlaamse Radio- en Televisieomroeporganisatie (VRT).

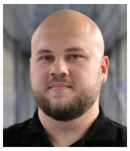
References

1. SMPTE, ST 259:2008, "SDTV - Digital Signal/Data — Serial Digital Interface," in ST 259:2008 , vol., no., pp.1-18, 29 Jan. 2008, doi: 10.5594/SMPTE.ST259.2008.
2. Audio Engineering Society (AES), AES67-2018: AES standard for audio applications of networks - High-performance streaming audio-over-IP interoperability," [Online]. Available: https://www.aes.org/publications/standards/search.cfm?docID=96. [Accessed 04 October 2023].
3. SMPTE, OV 2110-0:2018, "Professional Media over Managed IP Networks Roadmap for the 2110 Document Suite," in OV 2110-0:2018, pp.1-4, 24 Jan. 2019, doi: 10.5594/SMPTE.OV2110-0.2018.
4. SMPTE, ST 2110-21:2017, "Professional Media Over Managed IP Networks: Traffic Shaping and Delivery Timing for Video," in ST 2110-21:2017 , vol., no., pp.1-17, 27 Nov. 2017, doi: 10.5594/SMPTE.ST2110-21.2017.
5. SMPTE, RP 2110-25:2023, "Professional Media over Managed IP Networks: Measurement Practices," in RP 2110-25:2023, pp. 1-20, 6 July 2023, doi: 10.5594/SMPTE.RP2110-25.2023.
6. Institute of Electrical and Electronics Engineers (IEEE) "IEEE Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems," in IEEE Std 1588-2019 (Revision ofIEEE Std 1588-2008), pp. 1-499, 16 June 2020, doi: 10.1109/IEEESTD.2020.9120376..
7. T. Kernen, W. Vermost, I. Kostiukevych and P. Kondratenko, "An open platform for media over IP network load testing with off-the-shelf hardware," *Proc. NAB BEITC*, Las Vegas, 2023.

1545-0279/24©2024SMPTE

8. RFC 6985 - IMIX Genome: Specification of Variable Packet Sizes for Additional Testing," IETF, July 2013. [Online]. Available: https://datatracker.ietf.org/doc/rfc6985. [Accessed 04 October 2023].
9. "Elephant flow. (2023, August 25). In Wikipedia.," Wikipedia, 25 August 2024. [Online]. Available: https://en.wikipedia.org/wiki/Elephant_flow. [Accessed 14 December 2023].
10. T. Kernen and W. Vermost, "The art of conforming to SMPTE 2110-21 traffic model," *Proc. NAB BEITC*, Las Vegas, NV, 2018.
11. RFC 3376 - Internet Group Management Protocol, Version 3," IETF, October 2002. [Online]. Available: https://datatracker.ietf.org/doc/rfc3376/. [Accessed 04 October 2023].
12. RFC 7761 - Protocol Independent Multicast - Sparse Mode (PIM-SM): Protocol Specification (Revised)," IETF, March 2016. [Online]. Available: https://datatracker.ietf.org/doc/rfc7761/. [Accessed 04 October 2023].
13. Switched Port Analyzer (SPAN)," Cisco, 13 February 2020. [Online]. Available: https://learningnetwork.cisco.com/s/article/span-rspan-erspan. [Accessed 04 October 2023].
14. TAP Aggregation," Arista Networks, 2016. [Online]. Available: https://www.arista.com/en/solutions/tap-aggregation-with-danz. [Accessed 04 October 2023].
15. RFC 3176 - InMon Corporation's sFlow: A Method for Monitoring Traffic in Switched and Routed Networks," IETF, September 2001. [Online]. Available: https://datatracker.ietf.org/doc/rfc3176/. [Accessed 04 October 2023].
16. I. Kostiukevych, W. Vermost and P. Ferreira, "Analyzing SMPTE ST 2110 Streams Using EBU's Open-Source Software," *SMPTE Mot. Imag. J.*, 128 (4): 1-6, May 2019, doi: 10.5594/JMI.2019.2899712.
17. sFlow Version 5 Documentation, InMon Corp., July 2004. [Online]. Available: https://sflow.org/sflow_version_5.txt. [Accessed 04 October 2023].
18. Wireshark User's Guide, Wireshark Foundation, [Online]. Available: https://www.wireshark.org/docs/wsug_html/. [Accessed 04 October 2023].
19. sFlow Agents, InMon Corp, [Online]. Available: https://sflow-rt.com/agents.php. [Accessed 04 October 2023].
20. Tcpdump man page, The Tcpdump Group, 12 March 2023. [Online]. Available: https://www.tcpdump.org/manpages/tcpdump.1.html. [Accessed 04 October 2023].
21. I. Kostiukevych and W. Vermost, "High-Precision Capturing and Measuring of ST 2110 Streams Using Commodity IT Equipment," *Proc. NAB BEITC*, Las Vegas, NV, 2019.
22. RFC Draft - PCAP Capture File Format, IETF, 23 July 2023. [Online]. Available: https://datatracker.ietf.org/doc/html/draft-ietf-opsawg-pcap-03. [Accessed 04 October 2023].
23. M. Chiesa, G. Kindler and M. Schapira, "Traffic Engineering With Equal-Cost-MultiPath: An Algorithmic Perspective," in IEEE/ACM Transactions on Networking, vol. 25, no. 2, pp. 779-792, April 2017, doi: 10.1109/TNET.2016.2614247.
24. 7060X and 7260X Series 10/25/40/50/100G Data Center Switches Datasheet," Arista Networks, 15 March 2021. [Online]. Available: https://www.arista.com/assets/data/pdf/Datasheets/7060X_7260X_DS.pdf. [Accessed 04 October 2023].
25. Arista 7280R Series Data Center Switch Router Datasheet, Arista Networks, 20 June 2023. [Online]. Available: https://www.arista.com/assets/data/pdf/Datasheets/7280R-DataSheet.pdf. [Accessed 04 October 2023].
26. NVIDIA SN2010 Ethernet Switch for Hyperconverged Infrastructures Datasheet, NVIDIA, 2020. [Online]. Available: https://network.nvidia.com/files/doc-2020/pb-sn2010.pdf. [Accessed 04 October 2023].
27. "Imagine Communications Selenio Network Processor IP Media Processing Platform Datasheet," *Imagine Communications*, 2022. [Online]. Available: https://imaginecommunications.com/product/selenio-network-processor-snp. [Accessed 04 October 2023].
28. Bridge Technologies VB440 IP Probe Datasheet, Bridge Technologies, 2023. [Online]. Available: https://bridgetech.tv/products/vb440/?tab=Tech_specs. [Accessed 04 October 2023].
29. IMS - LANTIME M1000: Modular Synchronization Server for Communication Networks Datasheet, MEINBERG Funkuhren GmbH & Co. KG, 08 September 2023. [Online]. Available: https://www.meinbergglobal.com/download/docs/shortinfo/english/info_ims-lantime-m1000-telecom.pdf. [Accessed 14 October 2023].
30. SMPTE, ST 2110-20:2017, "Professional Media Over Managed IP Networks: Uncompressed Active Video," ST 2110-20:2017 , vol., no., pp.1-22, 27 Nov. 2017, doi: 10.5594/SMPTE.ST2110-20.2017.
31. SMPTE, ST 2110-22:2022, "Professional Media Over Managed IP Networks: Constant Bit-Rate Compressed Video," ST 2110-22:2022 , vol., no., pp.1-6, 12 Sept. 2022, doi: 10.5594/SMPTE.ST2110-22.2022.
32. Joint Picture Experts Group (JPEG) JPEG XS, ISO/IEC 21122 suite of standards, ISO 35.040.30 working group, https://www.iso.org/ics/35.040.30/x/.
33. SMPTE, ST 2110-30:2017, "Professional Media Over Managed IP Networks: PCM Digital Audio," ST 2110-30:2017, pp. 1-9,27 Nov. 2017, doi: 10.5594/SMPTE.ST2110-30.2017.
34. SMPTE, ST 2110-31:2018, "Professional Media Over Managed IP Networks: AES3 Transparent Transport," ST 2110-31:2018, pp. 1-12, 31 Aug. 2018, doi: 10.5594/SMPTE.ST2110-31.2018.
35. SMPTE, ST 2110-40:2018, "Professional Media Over Managed IP Networks: SMPTE ST 291-1 Ancillary Data," ST 2110-40:2018, pp. 1-8, 25 April 2018, doi: 10.5594/SMPTE.ST2110-40.2018.
36. InfluxDB v2.X Documentation, InfluxData, Inc., 2023. [Online]. Available: https://docs.influxdata.com/influxdb/v2/. [Accessed 04 October 2023].
37. Prometheus Data Model documentation, The Linux Foundation, 2023. [Online]. Available: https://prometheus.io/docs/concepts/data_model/. [Accessed 04 October 2023].
38. sFlow-RT, InMon Corp., 2023. [Online]. Available: https://sflow-rt.com/index.php. [Accessed 04 October 2023].
39. sFlow-RT Prometheus Exporter, sFlow Corp., 23 April 2019. [Online]. Available: https://blog.sflow.com/2019/04/prometheus-exporter.html. [Accessed 04 October 2023].
40. sFlow-RT InfluxDB v2.X integration, InMon Corp., 9 March 2021. [Online]. Available: https://blog.sflow.com/2021/03/influxdb-20-released.html. [Accessed 04 October 2023].
41. sFlow-RT Flow Definition API, sFlow Corp., 2023. [Online]. Available: https://sflow-rt.com/define_flow.php. [Accessed 04 October 2023].
42. Docker Compose documentation, Docker Inc., 2023. [Online]. Available: https://docs.docker.com/compose/. [Accessed 14 October 2023].
43. Internet Engineering Task Force (IETF), RFC 4566 - SDP: Session Description Protocol, IETF, July 2006. [Online]. Available: https://datatracker.ietf.org/doc/html/rfc4566. [Accessed 04 October 2023].
44. Standard sFlow v5 Data Formats, InMon Corp., [Online]. Available: https://sflow.org/SFLOW-STRUCTS5.txt. [Accessed 04 October 2023].
45. Catalog of structure numbers in sFlow v5 specification and extensions, inMon Corp., 2023. [Online]. Available: https://sflow.org/developers/structures.php. [Accessed 04 October 2023].

About the Authors



Ievgen Kostiukevych is the team leader for media over IP and cloud technologies at the European Broadcasting Union and EBU Academy Live IP faculty member. He spent over a decade in media over IP design, integrations, and solutions architectures before joining the EBU Technology & Innovation team.

Thomas Kernen is a principal architect at NVIDIA, Zurich, Switzerland. His main area of focus is defining architectures for transforming the broadcast industry into an all-IP infrastructure. He is a SMPTE Fellow, a standards director, a former chair of SMPTE's 32NF technology committee.

Willem Vermost serves as head of media production facilities at VRT. Prior to this role, he was the topic lead on the transition to IP-based studios at the European Broadcasting Union (EBU). He received a master's degree in electronic engineering and in applied computer science.

Pavlo Kondratenko is a project manager (media production over IP networks) at EBU Technology & Innovation. His background is in network engineering. He is a document editor of PICS for SMPTE ST 2110 standards suite.

Presented at the 2023 Media Technology Summit. Hollywood, CA, 16-19 October 2023.
Copyright © 2024 by SMPTE.

DOI: 10.5594/JMI.2024/LHXY3357
Date of publication: 24 January 2024

KEYWORDS VERSATILE VIDEO CODING // H266 // VIDEO COMPRESSION // VIDEO STREAMING // VIDEO CODECS // GUIDELINES

Configuring Versatile Video Coding:

technical guidelines for broadcast and streaming applications

By Lukasz Litwic Ericsson, Dmytro Rusanovskyy, Sean McCarthy, and Alan Stein

Abstract
Versatile Video Coding (VVC or H.266) is the latest video coding standard jointly developed by the International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) Motion Picture Experts Group (MPEG) and International Telecommunication Union-Telecommunication (ITU-T) Video Coding Experts Group (VCEG). With best-in-class compression performance, VVC can enhance existing applications and enable new services. As the first VVC implementations enter the market, several application-oriented standards-developing organizations and industry fora are defining VVC-based profiles and corresponding receiver capabilities. However, these specifications don't typically prescribe how a service is realized and the impact of the codec's operational parameters on delivered compression performance. To this end, the Media Coding Industry Forum has developed VVC technical guidelines. These guidelines will serve as a reference for VVC configuration choices to address operational, interoperability, and regulatory needs while achieving optimal compression performance. This paper presents an overview of the guidelines' scope, followed by a discussion of VVC configuration aspects, with focus on new features that are of utmost relevance to broadcast and streaming.

Versatile Video Coding (VVC) was standardized by ITU-T as Recommendation (Rec.) H.266 and in the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC) as International Standard 23090-3 (MPEG-I Part 3).[1] VVC is the latest generation of open video coding standards developed by the Joint Video Experts Team (JVET) of ISO/IEC JTC 1/SC 29 (known as Moving Picture Experts Group MPEG) and by ITU-T Question 6/16 of Study Group 16 (known as Video Coding Experts Group [VCEG]). The first release of the standard, published in July 2020, delivered a substantial increase in compression efficiency over its predecessor, High Efficiency Video Coding (HEVC)/H.265, achieving around 50% bitrate reduction for the same video quality. VVC also improves on HEVC by including a combination of efficient coding tools designed from the start to support a wider range of video content properties, including high dynamic range (HDR), wide color gamut (WCG), and computer-generated imagery for gaming and remote screen content sharing. VVC's efficient coding tools and rich functionality target efficient delivery of established and emerging video formats such as ultra-high-definition (UHD) with 4K (UHDTV-1) and 8K (UHDTV-2) resolutions, virtual reality (VR) 360⁰ video, and includes built-in scalability support already at a single-layer bitstream level. Additional versatility can be achieved by metadata signaled in supplemental enhancement information (SEI) messages specified in VVC itself and in VSEI (Versatile SEI messages for coded video bitstreams) (ITU-T Rec H.274 | ISO/IEC 23002-7[2]). Overall, VVC's coding tools, functionality, and metadata signaling capabilities make it the leading state-of-the-art, feature-complete video codec available.

Unlike HEVC, where new functionality and specialized coding tools were added incrementally over time (e.g., 4:2:2

1545-0279/24©2024SMPTE

and 4:4:4 chroma formats were added in HEVC edition 2, scalability support in HEVC edition 3, and screen-content coding in HEVC edition 4), VVC includes support for all this functionality already from the first release. Three sets of profiles were included to support a host of video applications:

- **Main 10** and **Main 10 4:4:4** profiles, where **Main 10** is the flagship profile expected to support a wide range of broadcast and streaming applications and includes support for 8- and 10-bit video, temporal sub-layer functionality, and spatial scalability through resolution change. **Main 10 4:4:4** includes support for additional chroma sampling formats: 4:2:2 YCbCr chroma typically used in video production and contribution applications and 4:4:4 RGB format for applications operating on content out of graphics frame buffers, e.g., remote desktop.

- **Multilayer Main 10** and **Multilayer Main 10 4:4:4** extend **Main 10** and **Main 10 4:4:4** profiles, respectively, with support for multi-layer bitstreams.

- **Main 10 Still Picture** and **Main 10 4:4:4 Still Picture** specialize **Main 10** and **Main 10 4:4:4 profiles** to support bitstreams comprising a single intra-coded picture.

The second edition of the VVC standard published in April 2022 added new operation range extension profiles supporting bit depths up to 12 bits for YCbCr chroma formats, intra-only profiles including up to 16 bits for RGB formats targeting high-fidelity content acquisition and studio production applications.

Following VVC finalization, VVC support was enabled in all relevant media transport and systems standards. Encapsulation in MPEG-2 TS for digital television broadcast services is specified in Rec. ITU-T H.222.0 | ISO/IEC 13818-1.[3] Encapsulation of VVC bitstreams in the MPEG ISO base media file format (ISOBMFF) is specified in ISO/IEC: 14496-15:2017.[4] VVC was also added to the MPEG Common Media Application Format (CMAF) including baseline VVC media profiles and multilayer VVC media profiles.[5] Main 10 profile is expected to provide the requisite compression performance and functionality for most typical broadcast and streaming applications. Such streamlining of available profiles compared with previous generations of codecs like AVC/H.264 and HEVC/H.265 is expected to have a positive impact on the cost of deployment and interoperability across the ecosystem. At the same time, since VVC includes a much richer set of coding tools and functionality than previous codecs, it would be beneficial to have an industry reference for VVC to guide configuration choices that address relevant user-profiles and operating points such as those defined in industry-leading application specifications. Such a reference for VVC could facilitate cross-industry harmonization and enhance interoperability across the broadcast and streaming ecosystem.

These VVC technical guidelines aim to provide information to facilitate incorporating VVC into application-oriented standards, media production workflows, and media distribution to consumers. In addition to covering best practices for VVC configuration in terms of functionality and compression performance, the guidelines aim to include up-to-date information on VVC operating bitrate ranges and provide information on the usage of VVC and SEI messages.

The Media Coding Industry Forum is an open industry forum with the purpose of furthering the adoption of MPEG standards, initially focusing on VVC, by establishing them as well-accepted and widely used standards for the benefit of consumers and industry.[6] With the strong industry VVC adoption, the MC-IF Interoperability Working Group has launched work on VVC technical guidelines with a focus on broadcast and streaming applications. These VVC technical guidelines aim to provide information to facilitate incorporating VVC into application-oriented standards, media production workflows, and media distribution to consumers. In addition to covering best practices for VVC configuration in terms of functionality and compression performance, the guidelines aim to include up-to-date information on VVC operating bitrate ranges and provide information on the usage of VVC and SEI messages.

1545-0279/24©2024SMPTE

This paper attempts to review the status of VVC deployment and adoption in application standards, introduces MC-IF VVC technical guidelines work, and provides an in-depth discussion on the performance or functionality of selected VVC configuration aspects. The authors of this paper represent companies who are members of the Media Coding Industry Forum and involved in work on MC-IF guidelines.

VVC Deployment Status

Thanks to strong industry interest, VVC has enjoyed a rapid roll-out across different parts of the entire end-to-end video ecosystem, with the first reports of optimized decoder implementation made during VVC standardization.[7,8]

Software Decoding

One of the key enablers of such a speedy implementation cycle is the relatively low complexity of the VVC decoding process, which is only about 1.5-2 times the complexity of an equivalent HEVC decoder.[9] Several publicly available software decoders and players have been released for various platforms: HD playback on Android[12,11,14] and iOS[10,12] mobile platforms, UHD-1 decoding on laptop/desktop Apple M1,[13] Intel i7 and Intel i9 processors[9] and UHD-2 decoding with AMD EPYC[9] and Intel Xeon including support for Multilayer Main 10 profile.[16] VVC was also reported to playback HD content in Edge, Firefox, and Chrome browsers using a WebAssembly build.[14,15]

Hardware Decoding

While software decoding enables early technology integration and launch of services, in the long term, deployments in broadcast and streaming ecosystems are expected to rely on hardware decoder capabilities, especially for UHD-1 and UHD-2-based video formats, including high frame rate (HFR) support (100/120 Hz frame rates). Availability of VVC decoder IP core for UHD-2 120,[17] SoC decoder for set-top boxed,[18] and TV chipsets with VVC support for UHD-1 and up to 4320p/120 were reported[19] while the first news of consumer TV sets supporting VVC were reported for 2023.[20]

Encoding

On the encoding side, several reports have been made since the launch of VVC. In October 2020, a publicly available optimized offline encoder was reported offering performance on par with the VVC reference software encoder (VTM).[8] Due to the inclusion of special encoder features such as adaptive QP, the implementation was demonstrated to outperform the reference VTM encoder in subjective verification tests against HEVC[21] while operating with 100x encoding speed that of the VTM. Several other offline encoders have subsequently been reported offering over 30% performance gains over HEVC[22,23] and have been integrated into cloud-based encoding,[24] transcoding,[25] and mobile over-the-top (OTT) services.[26,27] VVC encoding software with capabilities up to UHD-2 resolution,[28] including capabilities of real-time 4320p/30 encoding was reported.[29] Encoder vendors also reported compression performance gains obtained with VVC encoders computing resources comparable to those currently deployed for HEVC encoders. In terms of real-time UHD encoding deployed on a public cloud, VVC was reported to achieve between 15%-30% lower bitrate at the same quality as a commercial HEVC solution at 36% compute cost overhead.[30] Data from another vendor reported VVC real-time implementation to provide between 10%-15% bitrate reduction over HEVC while utilizing the same core coding engine as HEVC encoder[37] for UHD-1 content, and on average, 20% bitrate reduction for UHD-2 content, thus reducing UHD-2 delivery bitrates to fit DVB-T2 or ATSC-3.0 transmission systems at around 35 Mbits/s.[31] While these performance gains are reported lower than the 50% usually expected from a new generation of ITU-T and ISO/IEC coding standards, it is important to note that these came from early implementations, and encoder complexity in these cases was reported to be between 1x-1.5 times of that of equivalent HEVC encoder.

Interoperability

VVC interoperability with several key technologies in the broadcast and streaming ecosystem has also been demonstrated in end-to-end trials. Over-the-air (OTA) and OTT trials of UHD content with VVC demonstrating encapsulation with MPEG-2 over DVB-S2 and ISOBMFF delivered with DASH were reported in Ref. 37. Live UHD-1 VVC streaming with low latency CMAF packaging was also reported in Ref. 37. Interoperability with CMAF was also reported for Multilayer Main 10 decoder on UHD-2/UHD-1/HD content.[33]

Conformance Testing

JVET has developed a normative specification, Conformance testing for VVC as ITU-T Rec. H.266.1 and ISO/IEC 23090-15.[34] The specification includes a conformance bitstream test set comprising 280 bitstreams in over one hundred categories to cover all VVC profiles in version 1. DVB released VVC test content for verification and validation of DVB VVC profiles (see also DVB).[35] Commercial conformance and test streams for VVC were also developed already during the standardization phase and released shortly after standard completion.[36]

In summary, the deployment of VVC has a certain success for such a short term. The annual video developer survey report by Bitmovin, conducted with 424 respondents from over 80 countries,[38] indicated that 19% of the developer were using VVC in production for live coding and 15% for video on-demand coding, making VVC the third most reported codec in the survey with H.264/AVC indicated by 78% and 85% respondents and HEVC/H.265 indicated by 40% and 42% respondents, respectively. A comprehensive review of reported VVC deployments can be found in Ref. 39.

VVC Adoptions in Applications Standards

Video traffic has been growing exponentially. In mobile networks, the share of video traffic is estimated at 70% and is projected to increase to around 80% share of the traffic by the end of the decade.[40] This growth is partly

1545-0279/24©2024SMPTE

driven by the increasing consumer demand for content delivered in high-quality formats, which can either mean video delivery in new enhanced and immersive formats (e.g., with UHD-2 resolution) or a more extensive availability of UHD content delivered to customers. This has led several standards-developing organizations (SDO) across broadcast and streaming ecosystems to investigate the benefits of VVC for inclusion into their specification and systems. The Advanced Television Systems Committee (ATSC) has noted their intention to include VVC in the ATSC 3.0 standard. A recent report on ATSC 3.0 and Global Convergence mentions explicitly that "*ATSC is currently specifying Versatile Video Coding (VVC) for inclusion in the ATSC 3.0 suite of standards.*" [41]

Streaming industry fora like DASH-IF and CTA-Wave added VVC profiles into their guidelines and specifications.[42,43] Digital Video Broadcasting (DVB) and Sistema Brasileiro de Televisão Digital (Fórum SBTVD) already adopted VVC involving a formal process of verification of compression performance claims against commercial requirements. Although it has not begun formally incorporating new codes into standards, 3GPP SA43GPP SA4 has completed a feasibility study on codecs performance, where initial performance analysis of new codecs such as VVC was examined. The results of this study were published in 3GPP TR 26.955.[59] The following paragraphs give a brief overview based on publicly available reports of VVC verification and adoption status in these organizations.

DVB

Digital Video Broadcasting (DVB) started investigating new commercial requirements for new video codecs in 2020. DVB established several performance-related commercial requirements to be met by new codecs, including the ability to deliver UHD-2 video over legacy broadcast multiplexes at excellent quality; the ability to enable five UHD-1 services in a 40 Mbit/s DVB-T2 multiplex (compared to three UHD-1 services expected with HEVC); and provide at least 27% more efficient live broadcast encoding than HEVC, and over 30% performance gains for UHD-1 streaming use cases while maintaining performance gains for sub-UHD-1 resolutions.[44] DVB did not conduct internal tests but relied on external testing.[46] Within just a year, DVB verified VVC meeting its commercial requirement and adopted VVC into its codec toolbox, and released the Bluebook specification in February 2022.[45] Four VVC-based operating points for MPEG-2 transport stream and DVB-DASH delivery were defined. All four operating points include UHD with UHD-1 resolution and frame rates up to 60Hz. All operating points are also HDR-capable and progressive only. Maximum capability is defined for UHD-2 resolution with frame rates up to 120Hz. VVC support in DVB specification also includes new functionality to allow dynamic changes of bitstream resolution within a single-layer transport stream program. VVC also provides support for bitstream accessibility, such as composition of additional sign-language video within a main video bitstream.[46]

Fórum SBTVD

Sistema Brasileiro de Televisão Digital (aka Forum SBTVD) is the organization charged with creating broadcast and hybrid broadcast-broadband television standards in Brazil. SBTVD has been working toward a new standard for Brazil, TV 3.0. The process consisted of a call for proposals published in July 2020,[47] candidates' responses, testing and evaluation, and selection. Use cases and requirements in the Call for Proposal include delivery of UHD-2 OTA and UHD-2 via OTT, native HDR/WCG support, HFR, and a reduced-resolution portrait-mode closed-second video service, for sign language purposes. For video, VVC was selected as the video base layer codec, for both OTA and OTT delivery.

Remaining work (on the application layer) involves drafting specifications for audio and video coding and determining operating points based on subjective assessment of video coding quality. End-to-end demonstration of TV 3.0, which would include additional system components, i.e., physical and transport layer, is expected in August 2024.[48]

MC-IF VVC Technical Guidelines

The reported scale of investigations and evaluation processes conducted in applications standards SDOs points to the importance of assessing the true capabilities of new video coding technologies with respect to legacy codecs as well as alternative coding technologies. Once the new industry profiles are standardized, the evaluation of these new profiles against business needs falls on operators and service providers. A comprehensive set of advanced compression tools, rich functionality, flexibility of high-level bitstream operations, and interoperability with associated metadata contribute to the versatility of VVC and provide very flexible codec configuration options. To facilitate the industry adoption of VVC into standards, workflows, and services, MC-IF has launched work on VVC technical guidelines for broadcast and streaming applications. **Figure 1** shows a high-level diagram of the end-to-end broadcast and streaming ecosystem. The initial scope of the guidelines concerns primarily final emission to end-users. For this initial scope, focus is on the use of VVC Main 10 profile. However, other profiles may also be of interest, e.g., Multilayer Main 10, which was reported to be used in UHD-2 delivery scenarios, and Main 10 4:4:4 in case of primary distribution of video content with YCbCr 4:2:2 chroma format. The first release candidate of the guidelines was published in 2023.[6]

The guidelines aim to describe relevant VVC configuration options in reference to relevant industry VVC profiles, which span through the following areas:

- Examples of the application of VVC coding tools and VVC performance, including references to industry best practices, relevant external performance results, and studies.
- Examples of application of VVC functional features, including discussion of their benefits and interoperability aspects.

1545-0279/24©2024SMPTE

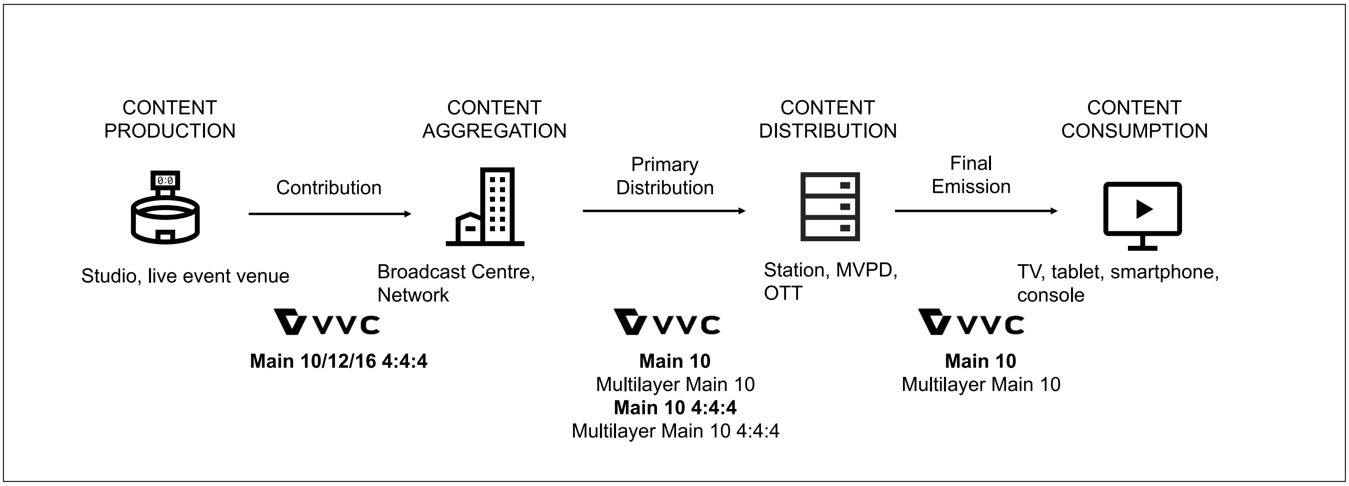


FIGURE 1. Simplified end-to-end broadcast and streaming ecosystem with content contribution, distribution, and final emission stages. VVC profiles relevant for respective parts of the ecosystem are highlighted.

- Examples of VVC usage with associated metadata.

In the following sections, we discuss selected examples across these categories, which are investigated for inclusion in the guidelines. The examples presented do not represent the full set of VVC options and capabilities. For a comprehensive description of VVC, please refer to the excellent overview.[49]

VVC Coding Tools and Performance: HDR Video

High Dynamic Range (HDR) video became a widespread video format in recent years. Multiple online streaming services such as Netflix, Amazon, Disney+ and others offer HDR content with hundreds of titles. HDR video services are also defined in the specifications of the TV broadcasts SDOs, such as ATSC, DVB and others. The development of HDR video formats in SMPTE, ITU-R and HEVC enabled rapid deployment of HDR video services. (HEVC Main 10 profile can compress 10-bit data and encapsulating HDR metadata into the coded bitstream. Although initially designed without HDR-specific coding tools, HEVC encoders can still be configured for efficient compression of HDR video data as well as signaling HDR-specific metadata. Operational practices of handling HDR video data and tool configurations in HEVC are documented in ITU-T Technical Reports (Supplements).[50,51]

The practices documented[50,51] are also relevant for VVC, but VVC takes HDR support to a different level. In VVC, all coding tools apply to both standard dynamic rate (SDR) and HDR video. The distinction between SDR and HDR is made only in high-level syntax. Specifically, the color primaries and optical-electronic transfer characteristics function that distinguish SDR, hybrid-log gamma HDR (HDR HLG), and perceptual quantizer HDR (HDR PQ) are signaled in the VVC bitstream using video usability information (VUI) metadata specified in VSEI. VSEI also specifies mastering display color volume (MDCV) and content light level indication (CLLI) SEI messages that provide metadata that are used to optimize the display of SDR and HDR-coded video on consumer displays. (See the section **Metadata for VVC** for more information.)

Although all VVC coding tools apply to both SDR and HDR, it is worth noting that some new VVC coding tools were initially studied in the context of HDR and later also to benefit SDR. In particular, the coding tools called luma mapping with chroma scaling (LMCS) and luma-intensity-based deblocking filtering adjustment aimed to facilitate compression of HDR video data. The luma mapping of the LMCS adjusts the dynamic range of the input video signal by redistributing the codewords across the dynamic range to improve compression efficiency. The chroma residual scaling of LMCS is utilized to compensate for the interaction between the luma signal and its corresponding chroma signals and reduce visual distortions in coded video with WCG representation (BT.2100).[52] For example, encoding algorithms to derive LMCS and luma intensity deblocking parameters are implemented in the VVC reference software (VTM) with specific optimizations for SDR, HDR HLG, and HDR PQ video data. For more details, see sections 3.7.2 and 3.7.3 in Ref. 53.

Objective quality metrics used during HEVC and VVC development for HDR video data have been documented in Ref. 54. It was determined that calculating the PSNR for the luma and chroma codewords, as typically for SDR video, is still appropriate for some HDR HLG video formats. For HDR PQ video, additional metrics were determined to be more suitable. These metrics include PSNRL100, wPSNR, and DE100, as described later and in more detail.[55]

- PSNRL100: This metric is calculated using the luminance values rather than the luma codewords. It is based on the CIELab color space representation of the input and output sample values of the codec.
- wPSNR: This is a PSNR-like metric calculated from the codewords that attempt to compensate for the more significant distribution of luma codewords to the darker regions by weighting of the codewords before calculating the PSNR.

- DE100: This is a metric based on the CIELab color space representation of the codec's input and output sample values. It is specifically targeted at chrominance fidelity.

The performance of VVC for coding HDR video data compared to HEVC, and other codecs has been assessed on multiple occasions. In document,[59] VVC performance assessment was conducted by the 3GPP SA4 working group. The coding performance of VVC for HDR PQ material was compared to the coding performance of HEVC, with both encoders using reference software implementations of JVET. For the set target bitrates, quality was assessed using objective metrics, namely wPSNR, PSNRL100, and DE100, as recommended in Ref. 54. It is reported that VVC provides 39% coding gain on average for the wPSNR metric, 35% for PSNRL100, and around 57% for the DE100 metric.

While objective results provide a good indication of performance, subjective assessment of the quality of coding performance is valuable not only to understand coding performance in relation to other codecs but also in relation to determining commercially relevant operating points. MPEG leads the industry best practice in this area and produces high-quality subjective verification tests for its codecs, performed by independent and qualified laboratories with the participation of naïve test subjects. Such tests follow established assessment protocols specified in ITU recommendations ITU-R BT.500[56] and ITU-T P.910[57] and define video quality evaluation in terms of Mean Opinion Score (MOS). MPEG subjective quality verification test results were reported in Ref. 32. The assessment included UHD-1 test sequences with HDR HLG and HDR PQ content. Reported MOS-over-bitrate results demonstrated an overall gain of about 49% for the HLG test sequences and about 52% for the PQ test sequences. **Figure 2** shows pooled MOS quality assessment data as a function of bitrate for tested VVC and HEVC codecs. The 11-grade MOS scale used in the test represents a full range of video quality impairments present in the tested bitstream, from a “0” score corresponding to “severely annoying (everywhere)” to “10” corresponding to “imperceptible.” “8” corresponds to “slightly perceptible (everywhere)” and can be associated with a broadcast quality. Plots in **Fig. 2** show that for VVC, averaged bitrates for UHD HDR content range from around 7 Mbits/s for PQ content used in the test to around 12.5 Mbits/s for HLG content used in the test. This constitutes a 50% bitrate reduction compared with HEVC at the same MOS quality achieved with average bitrates from around 15 Mbits/s for PQ content used in the test to around 27 Mbits/s for HLG content used in the test.

VVC Functional Features: Reference Picture Resampling

Delivery of video services by means of providing multiple renditions of content at varying resolutions and bitrates has been the backbone of adaptive streaming for broadband (DASH or HLS). Initially designed to target widely varying end-device display formats and variable available bandwidth, techniques such as “Per-title encoding” were introduced to dynamically select optimal resolutions (from a QoE perspective) for each VOD content. Decoding of short video segments with different bitrates and resolution could be achieved using closed group of pictures (GOP) coding, which inserts Instantaneous Decoder Refresh (IDR) at the start of each segment. In terms of compression performance, the closed GOP configuration is less efficient than the open GOP configuration typically used in broadcast. The coding penalty increases with a higher frequency of random-access points, i.e., shorter segments as IDR frames are less efficiently coded, which in effect causes bitrate spikes impacting rate control buffers and makes the penalty higher for low latency streaming. Previous attempts with seamless switching of resolutions with HEVC were reported[60] with the scalable profile of HEVC and more recently in[61] with the use of the Main 10 HEVC profile. However, the former approach could be viable only to applications supporting scalable HEVC profiles. At the same time, the interoperability tests with HEVC showed issues at switching since such functionality had not been tested at the time of deployment.

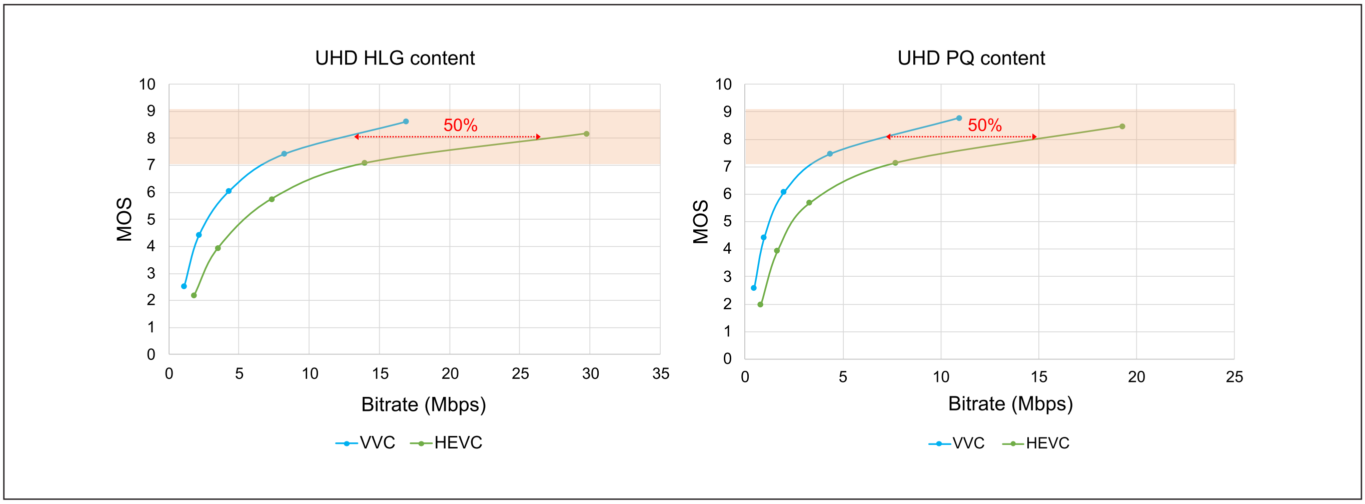


FIGURE 2. MOS quality over bitrate for UHD HLG and PQ content pooled over five test sequences in each category. Data to create these plots was used from Ref. 32.

1545-0279/24©2024SMPTE

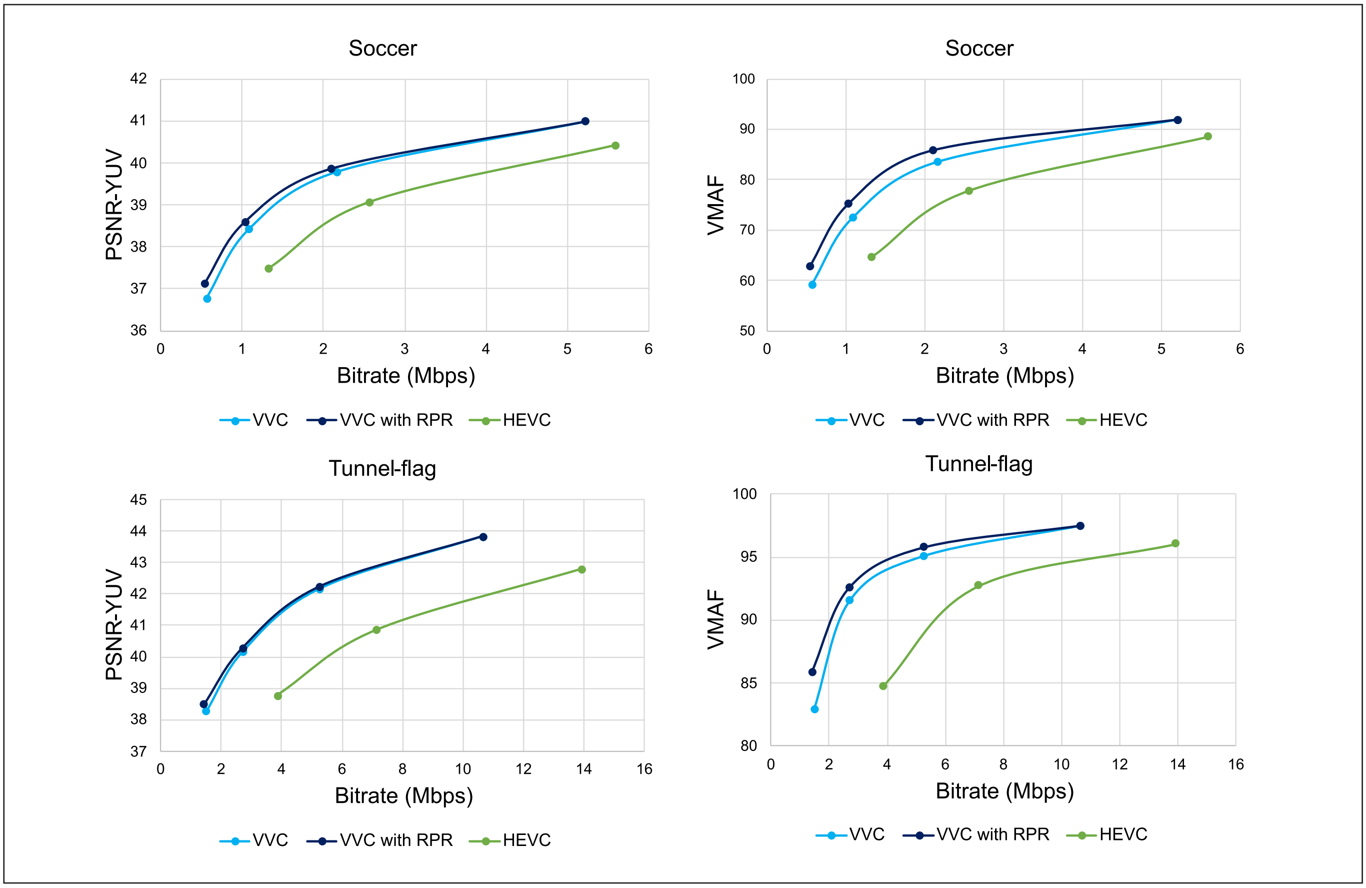


FIGURE 3. Rate-distortion plots for soccer and tunnel flag sequences for PSNR-YUV and VMAF metrics.

VVC includes a built-in spatial scalability functionality that allows changing resolution at any inter-coded picture. The mechanism uses reference picture resampling (RPR), which effectively adjusts the resolution of reference pictures used for prediction on the fly. This functionality is supported in the Main 10 profile while Multilayer profiles extend this through efficient high-level signaling. However, no additional inter-layer signal processing tools are required.[49] Skupin et al., investigated VVC open GOP resolution switching in the adaptive streaming use case and identified potential issues with prediction-error drift in random access skipped leading (RASL) pictures due to change of resolution at RAP (Intra picture), which precede RASL pictures in coding order.[62] It was, however, demonstrated that exercising certain constraints in the encoder mitigated most of the issues and resulted in compression gains up to 9% BD-rate over the typical closed GOP setting for low latency live streaming. Reference picture resampling functionality also enables changing resolution within a representation. While VVC allows high flexibility regarding scaling factors and frequency of resolution change at any picture, application specifications may provide some constraints on how often and which resolutions can be used by the codec. For example, the new revision of DVB TS 101 154,[45] which allows RPR use in VVC profiles, restricts downsampling scaling factors to 2/3 and 1/2. In contrast, new coded resolutions can be introduced at a minimum interval of 2 sec. However, different application standards may choose to adopt different sets of constraints for using RPR functionality.

To study the benefits of VVC RPR in the context of resolution change within a representation (e.g., as would be used in live broadcast), we used the VTM reference software implementation with RPR encoder functionality.[58] The encoding algorithm implemented in VTM aims to serve as an illustration of RPR functionality and is configured to limit resolution change at high bitrates while more likely to trigger downsampling at lower bitrates. Resolution change decisions can be made for each reordering segment of pictures (hierarchical B pictures structure) and can use scaling factors: 4/5, 2/3, and 1/2. We used sequences from 3GPP SA4 5G video test set for the test content.[59] Sequences used were from the SDR set of UHD-TV scenario (HDR results were reported in the previous section), for which overall VVC performance gains achieved with reference encoder VTM were reported as 37% BD-rate with PSNR-YUV, 33% BD-rate with PSNR-Y, 33% BD-rate with MS-SSIM and 38% BD-rate with VMAF metrics respectively. Details of these metrics, as used in the 3GPP 5G codec study, are detailed in section 5.5 of Ref. 59. In our test, we used four fixed QP settings (22, 27, 32, and 37) for VTM software as in the original study, plus QP 42 at the low end of the bitrate scale to align VVC and HEVC quality at the lowest bitrates.

Our tests showed substantial coding gains could be achieved for two out of eight sequences from the test set:

1545-0279/24©2024SMPTE

TABLE 1: BD-rate gains for VVC and VVC with RPR functionality over HEVC.

	BD-rate gains			
Test	**PSNR-Y [%]**	**PSNR-YUV [%]**	**MS-SSIM [%]**	**VMAF [%]**
Soccer				
VVC vs. HEVC	40.34%	39.27%	39.26%	40.23%
VVC with RPR vs. HEVC	44.10%	43.02%	43.67%	48.77%
Tunnel Flag				
VVC vs. HEVC	49.57%	52.64%	52.86%	55.19%
VVC with RPR vs. HEVC	51.08%	54.38%	57.21%	61.55%

Soccer[63] and Tunnel Flag.[64] For three of the sequences, either negligible gain or loss were observed. RPR functionality was not triggered, which is a correct behavior in case encoder does not predict content benefiting from change of resolution. However, this is still an interesting result since the sequences in the test set were not pre-selected towards showcasing RPR functionality, and RPR gains are expected to be content-dependent.

Results in **Table 1** show extra coding gain can be achieved over HEVC using the RPR coding tool in VVC. Coding gain was consistent for all reported metrics, with smaller difference reported by PSNR-based metrics and higher with MS-SSIM and VMAF. Looking at the rate-distortion plots in **Fig. 2**, we notice that coding gain is contributed through a higher quality metric value rather than shifted bitrates. Visual inspection showed that VVC is generally cleaner around the players than HEVC. At the same time, VVC with RPR looks more consistent than VVC and suffers less from visible artifacts in the background and textured areas like the one highlighted in **Fig. 3** for the Soccer sequence. **Figure 4** shows the RPR operation for the Tunnel Flag sequence at 5.2 Mbits/s. The sequence comprises two parts. The first part is to drive through a tunnel, which then cross-fades into a waving flag part. As shown in **Fig. 5**, RPR encoding algorithm utilizes three different scaling factors: 4/5 and 2/3 for the tunnel part and 1/2 scaling for the flag part, where the change occurs right after the cross-fade. While we did not conduct a detailed analysis of encoder complexity impact (encodings were run on a cluster with a mix of CPUs), ballpark results showed a decrease of about 15%-20% in computing resources when RPR was employed.

RPR is a new promising functionality offered by VVC that could benefit both resolution switching in ABR video delivery and introduce dynamic resolution changes within a bitstream or representation. In the latter case, it can either provide coding gain or optimize bitrate delivery for the same QoE, e.g., in statistical multiplexing by allowing greater flexibility in allocating bitrates across programs. RPR can also have an impact on CPU encoder complexity and therefore could contribute to reducing power consumption

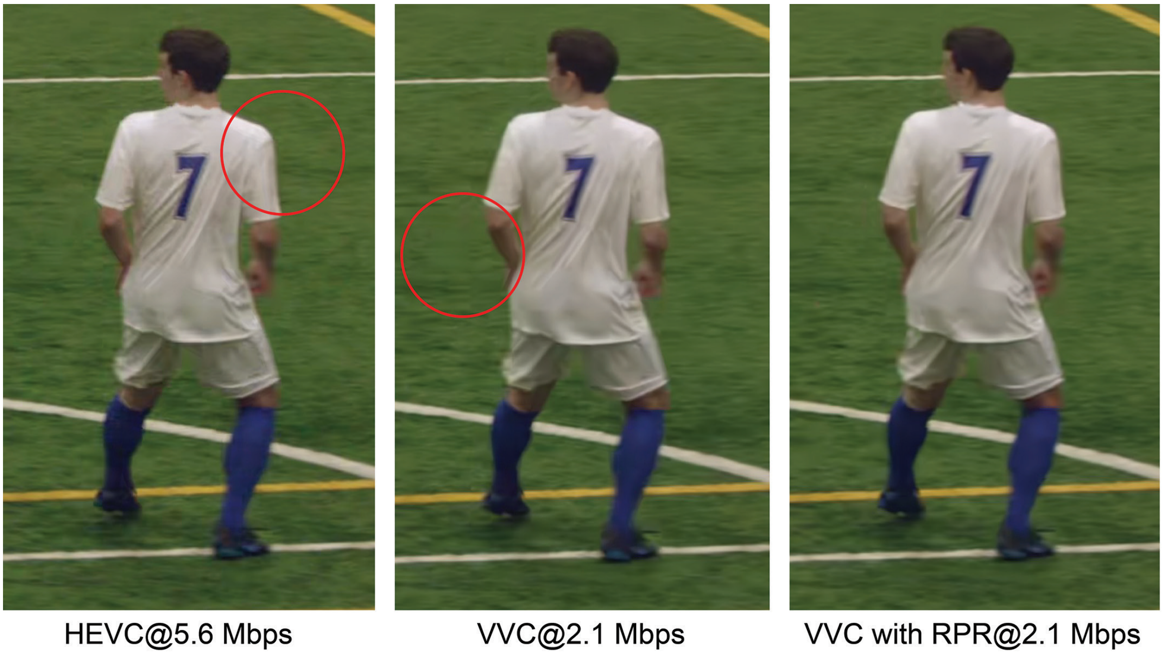


FIGURE 4. Screen shot from decoded sequence from HEVC bitstream (left), VVC bitstream (middle) and VVC bitstream with RPR (right). VVC with RPR decoded pictures were upscaled to the original UHD-1 resolution. Highlighted areas point to visible artifacts: halo around player's shoulder (left) and soft patch in the Astro turf (middle) sequence used in this test is from Ref. 63.

1545-0279/24©2024SMPTE

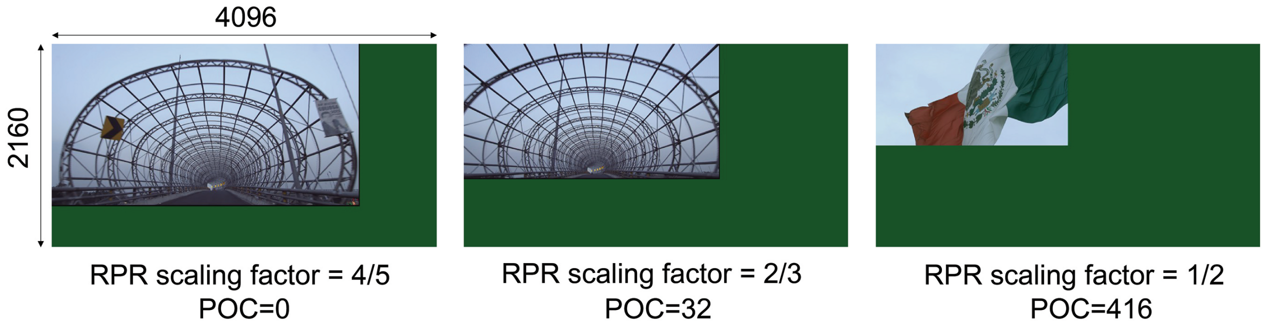


FIGURE 5. Visualization of RPR operation for Tunnel Flag sequence at QP 32 (5.2 Mbits/s). 4096 x 2160 (Digital Cinema Initiatives UHD-2 resolution) is the original resolution of the video sequence, which is represented by the green-padded area. Images represent decoded pictures resolution with POC (picture order count) representing timings where changes of resolution occurred. Sequence used in this test is from Ref. 64.

in the headend. The functionality is supported by some operational constraints in the DVB codec specification,[45] and scalability is also considered for SBTVD TV 3.0 specification requirements.[47]

Metadata for VVC

With a few notable exceptions, the VVC standard only specifies the syntax, semantics, and decoding processes required for conforming video decoders. Information about how video should be post-processed, displayed, or otherwise used is specified mainly in the VSEI standard. VSEI identifies two kinds of metadata: video useability information (VUI) and SEI messages.

VUI parameters provide information for the correct display of coded video. For example, VUI parameters indicate that the coded video is progressive or interlaced; SDR, HDR HLG, or HDR PQ; that the source video has a particular aspect ratio; and any other parameters necessary to correctly interpret the coded video.

SEI messages provide additional information that can assist decoders, displays, and other video receivers in performing as desired by the content producer. For example, the display of omnidirectional 360^{0}-video can be assisted by signaling one of the omnidirectional video-specific SEI messages in the VVC bitstream. As another example, the annotated regions SEI message can be used to assist video analysis applications by signaling the size and location of objects identified in a video.

SINCE ITS FINALIZATION IN 2020, **VVC HAS ENJOYED A STEADY ADOPTION PROGRESS** ACROSS VARIOUS IMPLEMENTATIONS SPANNING THE WHOLE END-TO-END CHAIN.

As noted previously, several SEI messages were originally developed in the context of HDR and have since become incorporated into application standards. The mastering display color volume (MDCV) SEI message provides information about how the source video looked to creative professionals when the content was created (i.e., mastered). Consumer displays can use the information in the MDCV SEI message to adjust how coded video is displayed to match the source look as closely as possible. The content light level information (CLLI) SEI message indicates how bright the source video was when mastered. Information in the CLLI SEI message can also be used to optimize the look of displayed video and has the potential to enable consumer displays to improve power management. Other SEI messages, including the content color volume (CCV), ambient viewing environment (AVE), and alternative transfer characteristics (ACT) SEI messages, can also enable consumer displays to process coded video content to provide better consumer experiences. It is worth noting that although many of the mentioned SEI messages were developed initially for HDR, they have become increasingly applicable to SDR video data as the brightness of SDR consumer displays has increased.

The film grain characteristics (FGC) SEI message has also become increasingly significant with the increase of film grain synthesis (FGS) in high-value streaming services.[65] The FGS SEI message signals the statistical characteristics of spatial, temporal grain added to the decoded video. There are two main use cases—one old and one new—for which adding grain to decoded video can be beneficial. The first and original use case was to replace actual film grain that had been filtered out of source video before encoding to save bandwidth. The newer use case is to subtly mask compression artifacts (even for source video that did not have original film grain) to reduce bitrate and improve perceived sharpness. A Technical Report on film grain technologies is currently being developed in ITU-T and ISO/IEC (as ISO/IEC 23009-9). Software illustrating the use of the FGS SEI message is available in the AVC, HEVC, and VVC reference software. Open-source film grain synthesis implementations are available in GitHub repositories.[67, 68]

Although not yet finalized, a pair of neural-network post-filter (NNPF) SEI messages[69] have attracted significant attention by enabling the use of neural networks for post-processing operations such as super-resolution, frame rate upsampling, chroma format conversion, and colorization. The NNPF characteristics (NNPFC) SEI message signals neural-network parameters/weights (contained in an ISO/IEC 15938-17 bitstream[70] or identified by a universal resource identifier[71]) and additional information needed for a receiver to determine if it can implement the indicated neural network. Several neural networks can be signaled using multiple NNPFC SEI messages to support different receiver capabilities and post-processing operations. Specific neural networks are invoked from the available neural networks using an NNPF activation (NNPFA) SEI message. The NNPFC and NNPFA SEI were planned to be finalized in VSEI in the second half of 2023.

Conclusion

Since its finalization in 2020, VVC has enjoyed a steady adoption progress across various implementations spanning the whole end-to-end chain. Also, application layer SDOs are looking into advancements in video codecs to further improve penetration of existing services while creating new operating points for delivering enhanced video services with formats such as UHD-2. The first adoption of VVC into application specifications for broadcast and streaming, MC-IF developed VVC technical guidelines. The guidelines are aimed to cover VVC configuration aspects, including its rich combination of coding tools, functionality, and associated metadata, to facilitate both VVC adoption and interoperability across products and services. In this paper, we provided an overview of VVC's deployment status across implementation and standards, discussed the scope of the launched work in MC-IF on VVC technical guidelines, and provided selected examples of VVC configuration aspects that demonstrate VVC superiority in terms of coding performance and functionality against the needs of the broadcast and streaming industry, and together with other aspects of VVC included in the MC-IF VVC technical guidelines.

References

1. International Telecommunication Union-Telecommunication (ITU-T) | International Organization for Standardization/International Electrotechnical Commission (ISO/IEC), "Rec. ITU-T H.266 | ISO/IEC 23090-3 "Information technology Coded representation of immersive media - Part 3: Versatile Video Coding," 2022.
2. International Telecommunication Union-Telecommunication (ITU-T) | International Organization for Standardization/International Electrotechnical Commission (ISO/IEC), "Rec. ITU-T H.274 | ISO/IEC 23002-7, "Versatile supplemental enhancement information messages for coded video bitstreams," 2022.
3. International Telecommunication Union-Telecommunication (ITU-T) | International Organization for Standardization/International Electrotechnical Commission (ISO/IEC), "Rec. ITU-T H.222.0 | ISO/IEC 13818-1: "Information technology - Generic Coding of moving pictures and associated audio information: Systems," 2022.
4. International Organization for Standardization/International Electrotechnical Commission (ISO/IEC), "ISO/IEC: 14496-15:2017: "Information technology—Coding of audio-visual objects—Part 15: Carriage of network abstraction layer (NAL) unit structured video in ISO base media file format," 2022.
5. International Organization for Standardization/International Electrotechnical Commission (ISO/IEC), "ISO/IEC 23000-19:2019/FDIS: "Information technology—Multimedia application format (MPEG-A) -- Part 19: Common media application format (CMAF) for segmented media (2nd edition)"," 2022.
6. Media Coding Industry Forum. [Online]. Available: https://www.mc-if.org/broadcast-streaming-guidelines/
7. F. Bossen, "Performance of a Reasonably Fast VVC Software Decoder," *JVET-S0224*, July 2020.
8. A. Wieckowski et al., "Open Optimized VVC Encoder (VVenC) and Decoder (VvdeC) Implementations," *JVET-T0099*, Oct. 2020.
9. F. Bossen, K. Sühring, A. Wieckowski, S. Liu, S., "VVC Complexity and Software Implementation Analysis," *IEEE Trans. on Circ. and Syst. for Vid. Technol*., 31 (10):3765-3778, Oct. 2021,
10. Y. Li et al., "Performance of a VVC Software Decoder on Mobile Platform," *JVET-U0071*, Jan. 2021.
11. L. Yu et al., "VVC Software Decoder Implementation for Mobile Devices," *JVET-V0132*, April 2021.
12. W. L. Feng et al., "VVC Software Decoder for Mobile Platforms," J*VET-V0070*, Apr. 2021.
13. L. Li et al., "Update on a VVC Software Decoder, BVC, for Heterogeneous CPU Plus GPU Systems," *JVET-Y0054*, Jan. 2022.
14. A. Wieckowski et al., "Update on Open, Optimized VVC Implementations VVenC and VVdeC" JVET-AB0044, Oct. 2022.
15. Fraunhofer HHI, "VVdeC Web Player," [Online]. Available: https://github.com/fraunhoferhhi/vvdecWebPlayer, accessed Feb. 2023.
16. Spin Digital. [Online]. Available: "Spin Digital Announces a VVC 8K Decoder and Media Player," June 2021
17. Allegro DVT. [Online]. Available: https://www.allegrodvt.com/products/al-d320-decoder-ip/, accessed Feb. 2023.
18. Realtek. [Online]. Available: "Realtek Launches World's First 4K UHD Set-top Box SoC (RTD1319D) Supports VVC/H.266 Video Decoding, GPU with 10-bit Graphics, Multiple CAS, and HDMI 2.1a", 29 Aug.2022.
19. MediaTek. [Online]. Available: MediaTek | Pentonic 2000 | Flagship 8K TV SoC, accessed February 2023.
20. R. Larsen. [Online]. Available "Sony and Philips will be the first TV makers to use MediaTek Pentonic chips", "*FlatPanelsHD*," 9 Jan. 2023.
21. International Telecommunication Union-Telecommunication (ITU-T) | International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) - ITU-T | ISO/IEC Joint Video Experts Team (JVET), "VVC verification test report for UHD SDR video content," *JVET-T2020*, Oct. 2020.
22. J. Cui et al., "An Optimized VVC Encoder Implementation," JVET-V0127, Apr. 2021.
23. Dong, X., et al., "Update on the Progress of the Optimized VVC Encoder Implementation, Ali266" JVET-X0104, Oct. 2021.
24. Bitmovin. [Online]. Available:"Bitmovin to Unveil Next-Generation VOD Encoder at the 2022 NAB Show," press release, 30 Mar. 2022.
25. Tencent.[Online]. Available: "Tencent Cloud Becomes First Cloud Service Provider to Support H.266/VVC Standard," press release, 16 Jul. 2021.
26. Y. Jia et al., "Ali266 @ Youku: trial deployment of VVC for video streaming," *JVET-Y0122*, Jan. 2022.
27. Best Media Info. [Online]. Available: "MX Player becomes first OTT to deploy H.266, cuts down video streaming data consumption into half," *Best Media Info*, 16 June 2021
28. *TVNewsCheck*. [Online. Available: "IBC 2022: MainConcept Launches VVC/H.266 SDK Beta And Demos Cloud-Ready 8K Broadcast Technology," *TVNewsCheck*, 31 Aug. 2022.
29. S. Sanz-Rodriguez, M. Alvarez-Mesa, and Ch. Ching Chi, "A VVC/H.266 Real-time Software Encoder for UHD Live Video Applications," *JVET-AB0043*, Oct. 2022.
30. J. Le Tanou, "MediaKind Enables Live UHD OTT VVC Streaming." [Online]. Available: "https://www.mediakind.com/blog/mediakind-enables-live-uhd-ott-vvc-streaming-on-a-public-cloud/", 21 Sep. 2022.
31. T. Biatek, M. Abdoli, T. Guionnet, A. Nasrallah, and M. Raulet, "Future MPEG Standards VVC and EVC: 8K Broadcast Enabler," *IBC* 2020.
32. International Telecommunication Union-Telecommunication (ITU-T) | International Organization for Standardization/International Electrotechnical Commission (ISO/

IEC) - ITU-T | ISO/IEC Joint Video Experts Team (JVET), “VVC verification test report for high dynamic range video content,” *JVET-W2020*, July 2021.
33. Japan Broadcasting Corporation (NHK) Science and Technology Research Labs (STRL) NHK STRL, “CMAF Multilayer-compatible VVC Decoder.” Press Release, 24 May 2022. [Online]. Available: https://www.nhk.or.jp/strl/english/news/2022/5.html
34. International Telecommunication Union-Telecommunication (ITU-T) | International Organization for Standardization/International Electrotechnical Commission (ISO/IEC), “Rec. ITU-T H.266.1 | ISO/IEC 23090-15 “ Information technology — Coded representation of immersive media — Part 15: Conformance testing for versatile video coding,” 2022.
35. Digital Video Broadcasting (DVB), “VVC test content.” [Online]. Available: https://dvb.org/specifications/verification-validation/vvc-test-content/, accessed 2 February 2023.
36. Allegro DVT. [Online]. Available: H266 VVC Standard JVET Syntax Compliance Bitstreams Validation Decoder (allegrodvt.com), accessed February 2023.
37. W. Hamidouche et al., “Versatile Video Coding Standard: A Review from Coding Tools to Consumers Deployment,” *IEEE Consumer Electronics Magazine*, 11(5):10-24, Sep. 2022.
38. Bitmovin, “The 6th annual Bitmovin Video Developer Report,” Dec. 2022.
39. G. J. Sullivan, “Deployment status of the VVC standard,” *JVET-AC0021*, Jan. 2023.
40. Ericsson, “Ericsson Mobility Report, Nov. 22. [Online]. Available: https://www.ericsson.com/en/reports-and-papers/mobility-report
41. Advanced Television Systems Committee (ATSC), “ATSC Planning Team Report: ATSC 3.0 and Global Convergence,” Dec. 2022
42. DASH Industry Forum, “DASH-IF Interoperability Points; Part 7: Video,” May 2022.
43. Consumer Technology Association (CTA), “CTA Specification, Web Application Video Ecosystem – Content Specification, CTA-5001-D,” 2021.
44. J. Power, “New Video Codecs for DVB Services: Fast-forward to 2024...,” *DVB Scene*, Issue 58, Sep. 2021.
45. Digital Video Broadcasting (DVB), “Specification for the use of Video and Audio Coding in Broadcast and Broadband Applications,” *DVB Document A001 Rev.21*, Nov. 2022.
46. V. Drugeon, “New in the DVB codec toolbox: Versatile Video Coding (VVC),” *DVB Scene*, Issue 59, Mar. 2022.
47. SBTVD Forum: Brazilian Digital Terrestrial TV Forum. TV3.0 Call for proposals. [Online]. Available: https://forumsbtvd.org.br/wp-content/uploads/2020/07/SBTVDTV-3-0-CfP.pdf
48. SBTVD Forum: Brazilian Digital Terrestrial TV Forum. [Online]. Available: Fórum SBTVD | TV 3.0 Project (forumsbtvd.org.br), accessed Feb. 2023.
49. B. Bross et al., “Overview of the Versatile Video Coding (VVC) Standard and its Applications,” *IEEE Trans. on Circ. and Syst. for Video Technol.*, 31(10): 3736-3764, Oct. 2021.
50. International Telecommunication Union-Telecommunication (ITU-T), “H. Sup15: Conversion and coding practices for HDR/WCG Y’CbCr 4:2:0 video with PQ transfer characteristics,” 2017.
51. International Telecommunication Union-Telecommunication (ITU-T), “H. Sup18: Signalling, backward compatibility and display adaptation for HDR/WCG video coding,” 2017.
52. International Telecommunication Union-Radiocommunication (ITU-R), ITU-R Recommendation BT.2100: Image parameter values for high dynamic range television for use in production and international programme exchange,” 2020.
53. International Telecommunication Union-Telecommunication (ITU-T) | International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) - ITU-T | ISO/IEC Joint Video Experts Team (JVET), “Algorithm description for Versatile Video Coding and Test Model 18 (VTM18),” *JVET-AB2002*, Oct. 2022.
54. International Telecommunication Union-Telecommunication (ITU-T), “Technical Paper ITU-T HSTP-VID-WPOM (07/2020): Working practices using objective metrics for evaluation of video coding efficiency experiments,” 2020.
55. International Telecommunication Union-Telecommunication (ITU-T) | International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) - ITU-T | ISO/IEC Joint Video Experts Team (JVET), “VTM and HM common test conditions and evaluation procedures for HDR/WCG video,” *JVET-AC2011*, Jan. 2023.
56. International Telecommunication Union-Radiocommunication (ITU-R), “ITU-R Rec. BT.500: Methodologies for the subjective assessment of the quality of television images,” 2019.
57. International Telecommunication Union-Radiocommunication (ITU-R), “ITU-T Recommendation P.910: Subjective video quality assessment methods for multimedia applications,” July 2022.
58. K. Andersson et al., “GOP-based RPR encoder control,” *JVET-AB0080*, Oct. 2022.
59. 3rd Generation Partnership (3GPP), “TR: 26.955: Video codec characteristics for 5G-based services and applications,” June 2022.
60. Y. Yan, M.M. Hannuksela, H. Li, “Seamless Switching of H.265/HEVC-coded Dash Representations with Open GOP Prediction Structure,” Proc. 2015 *IEEE International Conference on Image Processing (ICIP)*, pp. 4121-4125. 2015.
61. X. Ducloux, J-L Diascorn, T. Fautier, “Exploring the Benefits of Dynamic Resolution Encoding and Support in DVB Standards,” *IBC*, 2022.
62. R. Skupin et al., “Open GOP Resolution Switching in HTTP Adaptive Streaming with VVC,” *Proc. 35th Picture Coding Symp.* (PCS), pp. 1-5. June 2021.
63. Cable Labs. “Soccer.” [Online]. Available: https://dash-large-files.akamaized.net/WAVE/3GPP/5GVideo/ReferenceSequences/Soccer/Soccer.json, accessed Feb. 2023.
64. Netflix. “Tunnel flag.” [Online]. Available: https://dash-large-files.akamaized.net/WAVE/3GPP/5GVideo/ReferenceSequences/Tunnel-Flag/Tunnel-Flag.json, accessed Feb. 2023.
65. Y. K Wang et al., “The High-Level Syntax of the Versatile Video Coding (VVC) Standard,” *IEEE Trans. on Circ. and Syst. for Video Technol.*, 31(10): 3779-3800, Oct. 2021.
66. A. Norkin, and N. Birkbeck, “Film Grain Synthesis for AV1 Video Codec,” *Proc. 2018 Data Compression Conference,* pp. 3-12, March 2018.
67. Ittiam Systems. “Film Grain Synthesis Library.” [Online]. Available: https://github.com/ittiam-systems/libfgs, accessed Mar. 2023.
68. InterDigital Inc., “VersatileFilmGrain,” [Online]. Available: https://github.com/InterDigitalInc/VersatileFilmGrain, accessed Feb. 2023.
69. S. McCarthy, S. Deshpande, M. M. Hannuksela, G.J Hendry, Sullivan, and Y. KWang, “Improvements Under Consideration for Neural Network Post Filter SEI Messages,” *JVET- AC2032*, Jan. 2023.
70. International Organization for Standardization/International Electrotechnical Commission (ISO/IEC), “ISO/IEC: 15938-17, Information technology – Multimedia content description interface – Part 17: Compression of neural networks for multimedia content description and analysis,” 2022.
71. Internet Engineering Task Force (IETF), “IETF Standard 66, Uniform Resource Identifiers (URI): Generic Syntax.” [Online]. Available: http://tools.ietf.org/html/rfc3986

About the Authors

Lukasz Litwic, PhD, is a research leader at Ericsson. He joined Ericsson in 2007, where he worked on various aspects of image processing and video compression research, which formed the foundation of Ericsson real-time broadcast encoding products.

Dmytro Rusanovskyy is a principal engineer at Qualcomm Inc., where he works on algorithms design and standardization in areas of video compression, streaming and broadcasting.

Sean McCarthy is director of video strategy and standards at Dolby, where he explores innovations, new use cases, and core technology standards that assist Dolby transform storytelling, and produce new experiences that unleash the potential of entertainment and communications.

Alan Stein leads InterDigital’s Visual Standards and Business Strategy team, where he manages a global team of senior technology experts who participate in major video standards bodies and important industry forums.

This paper first appeared in the Proceedings of the NAB 2023 Broadcast Engineering and Information Technology (BEIT) Conference. Reprinted here with permission of the author(s) and the National Association of Broadcasters, Washington, DC.

DOI: 10.5594/JMI.2024/MRLF3850
Date of publication: 24 January 2024

1545-0279/24©2024SMPTE

OFF-TOPIC: APPLICATIONS/PRACTICES

KEYWORDS FINAL SAMPLE // SOUND MITIGATION SPECIALIST // PRODUCTION SOUND MIXER/RECORDIST

Improvements to On-set Virtual Production Acoustics for Production Dialog

By Eric Rigney

Abstract
While capturing in-camera visual effects (ICVFX) or ‘final pixel’ remains a primary goal of On-Set Virtual Production (OSVP), “final sample,” recording usable performance dialog that carries over to final sound mix, is often its casualty. Unlike the sound damping qualities of traditional sound stages, volume stages can resemble the acoustical characteristics of an echo chamber. Production sound mixer/recordists’ traditional abatement methods can prove inadequate in OSVP due to concerns over blocking visually valuable environmental lighting reflections and space limitations restricting reverberation mitigation options. Automated Dialog Replacement (ADR) is a post-production solution of last resort due to budgetary, logistical, and creative costs. Zero-ADR (ZADR) is therefore as creatively and economically as valuable to production as is the goal of final pixel.

With education, productions employing a sound mitigation specialist with specific mitigation equipment, and early and consistent cooperative/collaborative inter-departmental communication can effectively mitigate reverberation and unwanted noise in a volumetric stage, allowing post-production sound editors to effectively process (clean) the OSVP recorded tracks, significantly minimizing ADR usage.

Final sample is the concept and practice of capturing dialog of a sufficient quality in a volume stage that, with some post-production processing and without the use of automated dialog replacement (ADR), carries through to the final sound mix. When considering on-set virtual production (OSVP), sound recording challenges are not often discussed, nor are the additional ADR costs to repair unusable OSVP dialog. This study describes how simple pre-production and production efforts spent toward guarding original dialog performance are offset by reductions in ADR costs.

Ironically, many volume stages are built inside “soundstages,” and few, if any, volumetric stages appear to have been designed with input from sound designers or acoustic consultants. For filmmakers appreciating that ADR cannot replace the emotional impact of a given performance, final sample, or zero-ADR (ZADR) should prove valuable. For filmmakers and producers for whom ADR quality dialog is an easily acceptable patch, the results of this study may be less valuable.

Walking on an empty LED volume stage, it is apparent that little if any consideration was given toward abating the acoustical challenges that the OSVP environment created. In pursuit of in-camera visual effects (ICVFX), panel manufacturers, stage designers and builders, and most virtual production facility owners and producers unwittingly induced significant acoustical issues for productions recording dialog. Light-emitting diode (LED) volume stages are expensive for facilities to construct and operate and expensive for productions to rent. Current research and development in LED panel and stage design focus on image quality in pursuit of ICVFX goals, not on guarding sound integrity. Until changes in LED volume panel and stage design and construction

1545-0279/24©2024SMPTE

effectively reduce sound reverberation, reflections, and unwanted noise captures, OSVP in its current state requires added costs of labor and equipment to mitigate its reverberation imposition or accept that OSVP productions will suffer inferior sound and the associated loss of emotional impact on its audiences, or both. As renowned production sound professional Patrushkha Mierzwa stated, “highly reverberant audio [in a volumetric stage] is not the sound department’s problem, it’s the production’s problem.” Ultimately, the acoustical challenges imposed on productions are created by the producers who choose to rent reverberant environments. OSVP stages differ in shape and size. Some offer ceilings, and others offer floors. In choosing a virtual production stage, productions may wish to consider OSVP stages that best serve the sound and the visual needs of specific scenes.

> Final sample is the concept and practice of capturing dialog of a sufficient quality in a volume stage that, with some post-production processing and without the use of automated dialog replacement (ADR), carries through to the final sound mix.

Shared responsibility: Productions typically rely on the sound department to resolve acoustical problems with little rehearsal. OSVP introduces new and significant acoustical issues that production sound mixers/recordists are inexperienced in resolving and still with little preparation time. Production sound mixers/recordists are typically not acoustic engineers or experienced acousticians. And a typical three-person sound team, fully engaged in its current responsibilities (slate synching, on-set communications, configuring and monitoring multiple recording and wireless technologies, preparing and ‘mic’ing’ talent, setting up for the shot, mixing, etc.), don’t have the training or resources to implement additional sound mitigation to a facility. More importantly, for those filmmakers who want the artistry of their production sound mixer/recordists to capture high-quality dialog that fits the scene, adding the mitigation responsibilities to the workload of production sound crews will come at the sacrifice of something else and artistic frustration. Without additional training, labor, equipment, and budgetary support, current production sound teams are preoccupied, ill-equipped, and ill-prepared to adequately mitigate noise and reverberation in LED volume stages. Sound reverberation and noise reduction in an OSVP is a production problem that requires support from all on-set production departments. Good sound is not the responsibility of the sound departments alone.

“All for one and one for all.” At a minimum, it helps when all stakeholders acknowledge the acoustic challenges that must be addressed. Until LED panel manufacturers and existing stage designs address reflection mitigation, productions are left to address the reverberation problem that the new and quickly adopted OSVP technology imposes. Initial test results indicate that with sound mitigation guidance from acoustic engineers, acousticians, or experienced OSVP sound mitigation specialists, on-set production departments working together from pre-production through post-production may adequately reduce the acoustical reverberation and other defects in a volume stage. Tackled collectively, mitigation of acoustic reverberation and noise before production can work toward achieving the goal of final sample, but only if production personnel learn how and are willing to execute what is advised. It is important to reaffirm to producers and directors that no single solution, technological or departmental, can *fully* address reverberation and noise in an LED volume stage. Similar to image quality, good sound is the responsibility of all production departments.

Pictorial Analogy: To understand what is being asked of production sound teams, here’s a contrived pictorial equivalent. Imagine if LED volumes were designed to always display “full white,” with no attenuation ability, “off” button, or ability to match the virtual environment? In addition to the hostile lighting circumstance, there are no mitigation apparatuses to block, attenuate, or redirect the light, nor the personnel needed to place such equipment. Imagine that regardless of the scene to be filmed, the director of photography (DP) and crew were expected to “make it work” without rehearsals. Imagine if DPs were expected

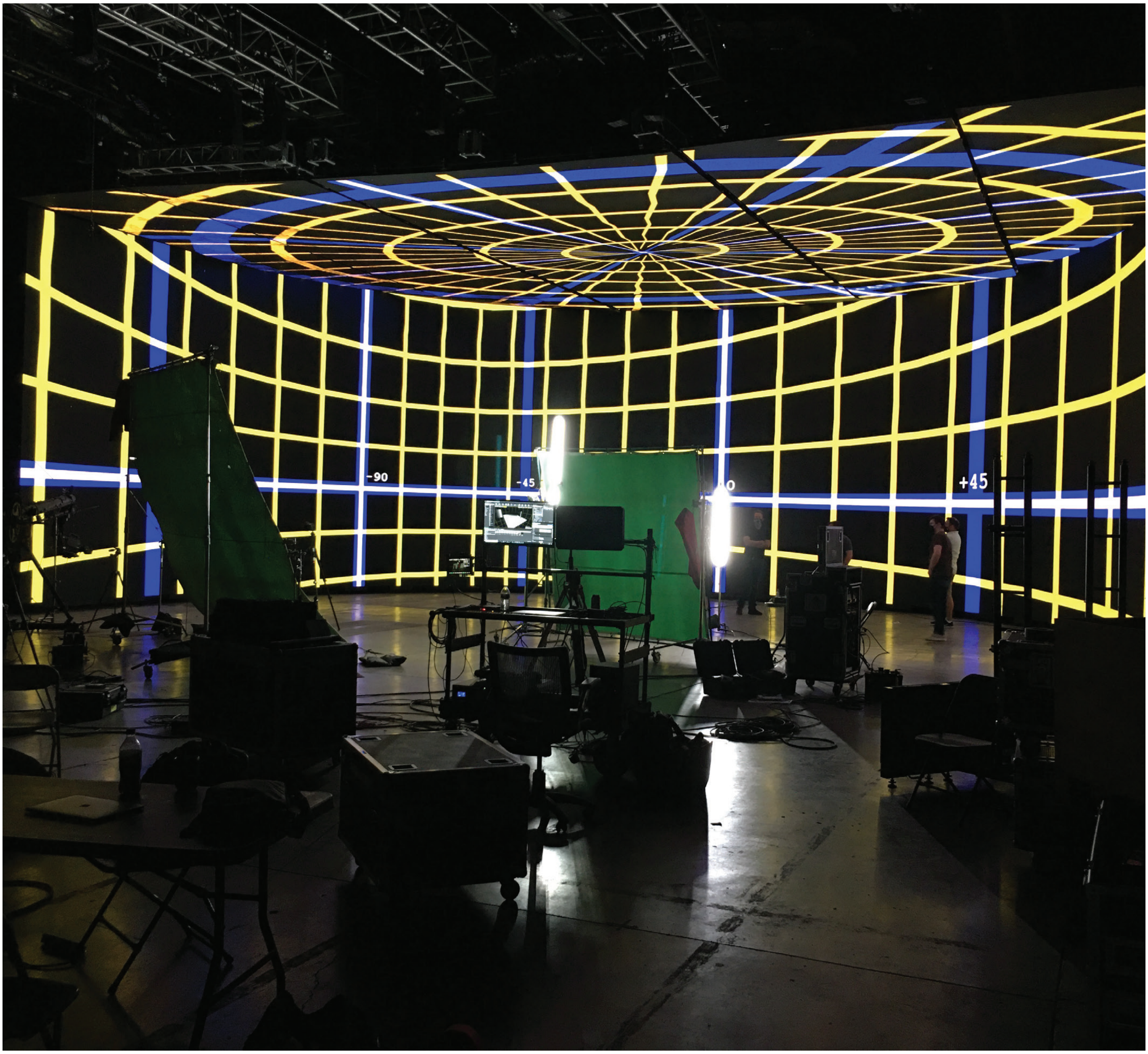


FIGURE 1. Semi-circular, 50-ft, LED volumetric stage displaying latitude and longitude lines developed by Paul Debevic to align wall and ceiling imagery.

to leave the art of production lighting to post-production visual effects and color correction artists. In today's LED volumetric stages, these are precisely the conditions under which productions expect their sound departments to perform, often intimating that automated dialog replacement (ADR) is an acceptable cure for combating high reverberance and other acoustical defects in OSVP.

Although sometimes necessary, ADR is emotionally inferior to production dialog. Of course, actors move around in virtual production volume stages just as they do on traditional sound stages, jostling props, sets, costumes, and other actors. Even with new mitigation attempts ADR will likely still be necessary, but to a significantly lesser degree.

Despite the inherent acoustical challenges and added costs of recording dialog in a volume stage, dedicated sound mitigation specialists such as acousticians and gaffer-like on-set sound mitigation labor could safeguard on-set dialog performances and stave off ADR.

The Experience: Walking inside a virtual production volume stage for the first time, the echoing of footsteps is clearly heard (**Fig. 1**). A whisper spoken directly against a wall on one side of the stage could be clearly heard from the other side of the stage when directly facing the opposite wall—the whispering gallery effect. A center-stage handclap tells the whole story. Feeling great empathy toward production sound mixer/recordists and post-production sound editors and mixers, Erik Weaver and Greg Ciaccio of the University of Southern California's (USC) Entertainment Technology Center (ETC) suggested we study the problem. The author's first call was to Sean Tajkowski (technical planning architect), who led him to Russ Berger (acoustical consultant), and so on.

This preliminary study demonstrates that acoustical responsibility lies across all OSVP departments, from LED panels and wall construction to final sound mix. Serving sound recording serves a production's audience, reducing costly and emotionally inferior ADR. As with any preliminary study, much was learned, with more to be studied and developed. To hear and see sound and video examples captured for this study, visit www.finalsample.com.

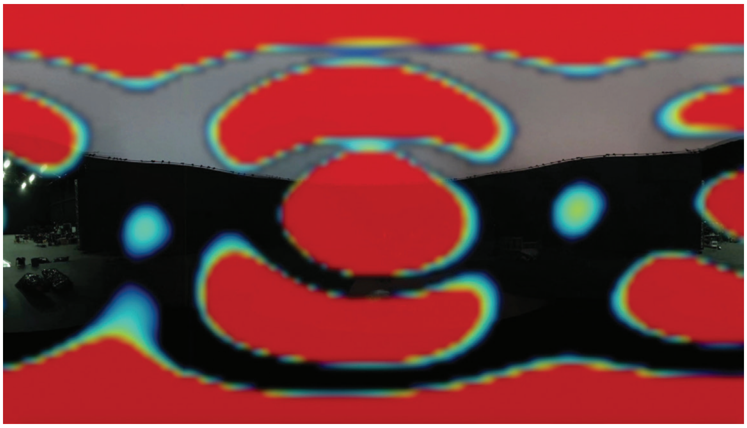


FIGURE 2. Acoustical camera view of balloon pop sonic instance recording and visually translating reverberation in an empty LED volumetric stage. The red color in the image indicates >75 dB between 2k-2.2 kHz.

Problem: Descriptions of Acoustical Challenges

Acoustical Reflection

Reverberation Time, RT60: A measure of reflection is the length of time for reverberation to decay by more than 60 dB. For most situations, inaudibility occurs before the full 60 dB is reached due to the limit of the signal to the ambient (background) noise. This is called the RT60 value. For dialog recording on a stage, the shorter the duration, the better.

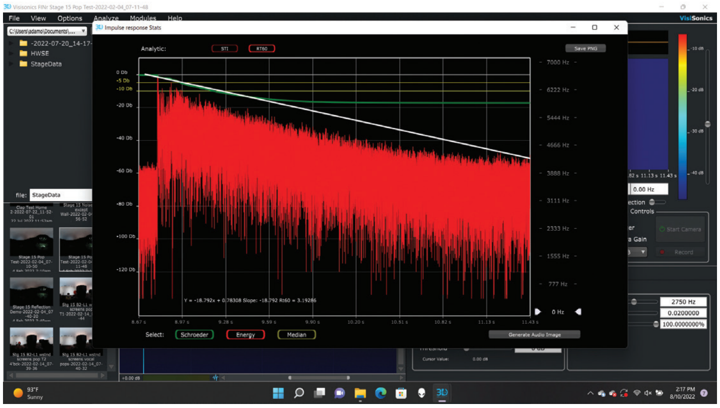
Post-Production Processing: De-reverberation software applications seem easily capable of correcting RT60 values of 1.6-1.8 sec or less.

Speech Transmission Index (STI): STI measures the intelligibility of speech in a given environment. Ambient noise and reverberation affect intelligibility. Mitigating either one helps increase STI; addressing both is best. Measured from 0.0 to 1.0, the higher the value, the better.

Volume Stage RT60 & STI Measurement: A balloon pop in the center of a large (80-ft diameter) empty LED volumetric stage, using an acoustical camera, measured an RT60 value of 3.2 sec. and an STI value of 0.51 (slightly above poor) (**Figs. 2, 3, and 4**), not quite the quality of a large cathedral ambiance, and great for recording music but not speech.

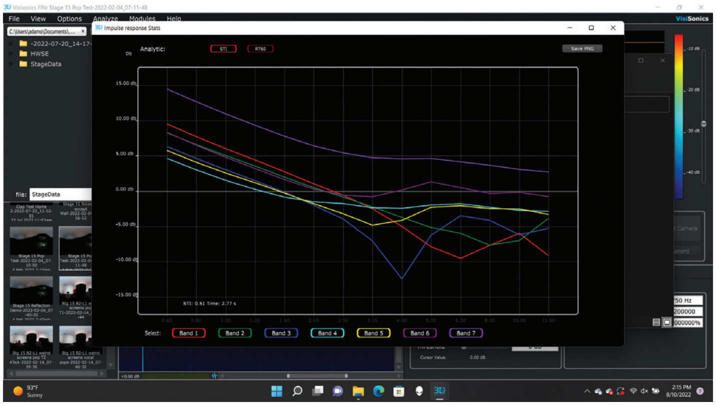
Light Emitting Diode (LED) Volume Stage Design and Construction: The shape, size, texture, and encompassing characteristics of LED volumetric stages affect reverberance. Walls vary in height. Shapes vary from a single flat wall to a "hockey stick" shape, to a three-sided open square, to a semi-circle, to "J-shaped," to an oval, to a closed circle, and everything in between. Ceilings, if installed, partially or completely cover the top. Some floors are polished cement, while others are raised, hollow, flat wooden structures. All these characteristics affect the reverberant quality in varying ways. Perpendicular, hard, and flat surfaces support reflective, reverberant acoustics.

Actor Blocking Locations: From where the actor speaks and in what direction affects the amount of reverberation captured in a dialog recording. Preliminary testing indicates that in an empty volume stage, standing center-stage and speaking toward a side wall generates the most reflection, as the sound bounces perpendicularly back and forth against the hard, flat surfaces of the parallel walls. Speaking toward the stage opening and/or with one's back near the wall (i.e., 6-8 ft) generates fewer reflections. Perhaps when facing the opening, the sound escapes through the opening, and less sound reflects. When speaking with one's back close to the wall, perhaps it is less reflective because sound has a greater distance to travel, or maybe the back wall acts like a reflective brake. More study is required.



Y = -18.792x + 0.78308 Slope: -18.792 Rt60 = 3.19286

FIGURE 3. Acoustical camera RT60 calculation (3.19 sec) of balloon pop sonic instance recorded in an empty LED volumetric stage without acoustical mitigation elements.



STI: 0.51 Time: 2.77 s

FIGURE 4. Acoustical camera STI calculation (0.51) of balloon pop sonic instance recorded in an empty LED volumetric stage without acoustical mitigation elements. Bands 1-7 depict standard STI frequencies of 125, 250, 500, 1K, 2K, 4K, 8K, respectively.

1545-0279/24©2024SMPTE

Acoustical Noise

Overview: While production sound mixer/recordists strive to capture clear, clean dialog, sometimes ambient sounds can be desirable, but not often. Ambient sounds for those whose origins are not captured simultaneously visually or that are seen but disrupt dialog are problematic. Production sound mixers/recordists strive to avoid recording them. After shooting, production sound mixers/recordists will often record background sounds for post-production sound editors to blend in and mask dialog recording or editing issues.

Noise Amplification Characteristic: The same characteristics of an LED volumetric stage that support a highly reverberant environment also amplify any noise captured in front of a stage opening. Due to the curved shape and smooth, hard surfaces of circular LED volume stages, they behave like parabolic microphones, efficiently capturing and increasing the noise. A creaking director's chair, keyboard strokes, a dropped water bottle, equipment cooling fans, page turns, footsteps, etc., are sounds that are easily captured and magnified within the sound recording area of an LED volume stage (**Figs. 5 and 6**).

Virtual Production Equipment Additions: In addition to traditional filmmaking technologies, on-set virtual production employs technologies specific to virtual production operations. Some of this equipment generates noise. For example, some stages set up the virtual art department's (VAD) technical control stations ('brain bar' or 'volume control') on the stage floor, in front of the stage opening. Some LED equipment is fan-cooled, and realtime processing equipment can also generate fan noise.

Virtual Production Personnel Additions: In addition to traditional filmmaking personnel, on-set virtual production stages and vendors employ personnel specific to virtual production operations: virtual art department artists (VAD), stage operations, camera technicians, etc., any one of which can unwittingly generate unwanted noise.

Solutions: Tools and Practices

Overview: This section describes how to collaboratively and cumulatively mitigate acoustical reverberation and unwanted noise in a volume stage. Collectively tackling acoustic reverberation and noise, facilities and productions can achieve *final sample*.

Virtual Production Equipment: OSVP introduces new technology and personnel. Could these elements be used to mitigate sound reflection and noise in an LED volume?

LED Panels: Panel construction differs. Some expose LED pixels, offering a slightly rougher texture, while others encase panel diodes with a clear cover, creating a smooth surface. The exposed LED version may be slightly less reflective, but both are hard, flat, and acoustically reflective. Future options discussed by panel manufacturers include installing shaped insulation between diodes or bursting micro laser holes through the clear-facing cover. A panel's cooling solution plays a role in acoustics. Noiseless heat sinks add weight, which is more of an issue for ceiling panels than wall or floor panels. For those panels that use fans, different fan designs emit different noise characteristics. Fortunately, panel fans are fitted on the backside of panels, directing most of the sound away from the interior of the volumetric stage. Providing noise-level specifications for panels may assist consumers in making acoustical choices.

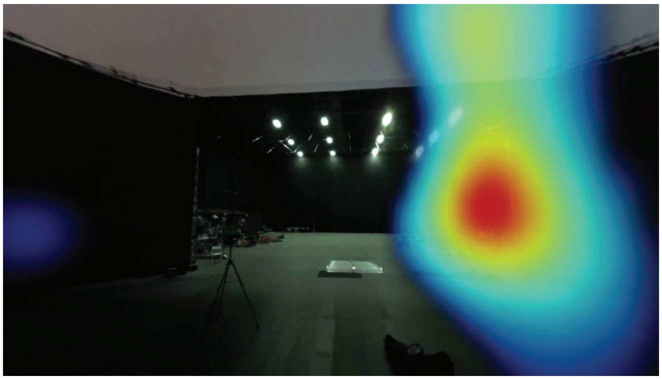


FIGURE 5. Acoustical camera capture of equipment noise recorded center stage of an empty LED volumetric stage without acoustical mitigation elements. The red color indicates >25-35kHz between 2-2.2 kHz.

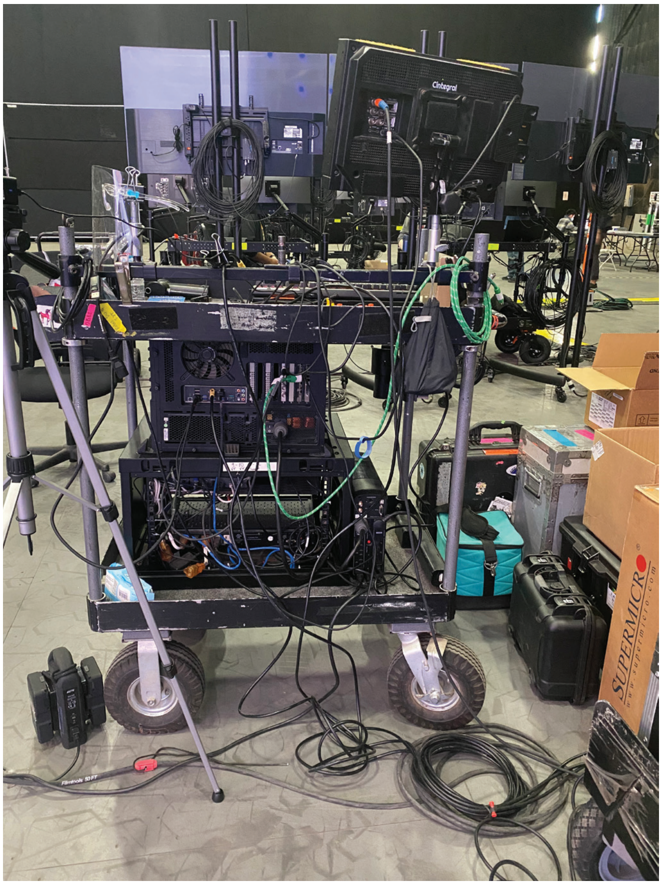


FIGURE 6. Digital Intermediate Technician's (DIT) equipment in front of the opening of an empty LED volumetric stage, the origin of the image in **Fig. 5**.

1545-0279/24©2024SMPTE

LED Walls: LED panels are configured into walls of different sizes and shapes: single flat wall, angled two-walled "hockey stick," open squares three flat walls, car processing, semi-circular (180^o), 'J-shaped' (270^o), and circular/oval (270^o-360^o). Some LED walls are installed in acoustically absorptive sound stages. Others are built in acoustically reflective warehouses. Each offers a different acoustical challenge.

a. *Pitch/Angle:* Acoustical engineer and designer Russ Berger and studio planning architect Sean Tajkowski suggest deflecting sound energy by pitching the walls back, like a control booth window in a music recording studio. With most panel LED pitches and resolutions supporting some off-axis angle viewing/recording, consider leaning (pitching) panels top-back 3-7°. Angled surfaces reflect sound upward and away from speaking actors and minimize bounce-back from opposite walls. Flat walls that are non-circular may be a cost-effective option. But singular, flat walls tend not to pose significant reverberance issues. For circular stages, pitching walls presents an engineering and maintenance challenge. A curved surface pitched back is a compound curve, creating a conical frustum (**Fig. 7**). Although possible, panels designed and manufactured to support a compound curve necessitate a custom build, which is expensive to purchase and maintain/replace.

b. *Environmental LED Lighting Alternatives:* An alternative to LED walls of flat panels is LED lighting that carries over virtualized environments, such as image-based lighting (IBL). One IBL concept uses tubes of light, mounted vertically or horizontally, through which sound can pass. The detail of imagery displayed by these specialized lights is significantly less than that of LED panels but offers a broader light spectrum of color beyond that of the limited color range of current red, green, blue (RGB) LED panels. Yet to be acoustically tested, intuitively, it would seem less acoustically reflective to talk into a wall of separated tubes than a solid wall of flat LED panels (**Fig. 8**).

c. *Acoustical Wall Curtains:* Lastly, a system of acoustic curtains covering unused portions of the walls could help abate acoustic reflection, deflecting/diffusing sound energy. A transparent curtain would allow some environmental light features to pass through. Curtains may also provide some protection to panels during set construction and dressing.

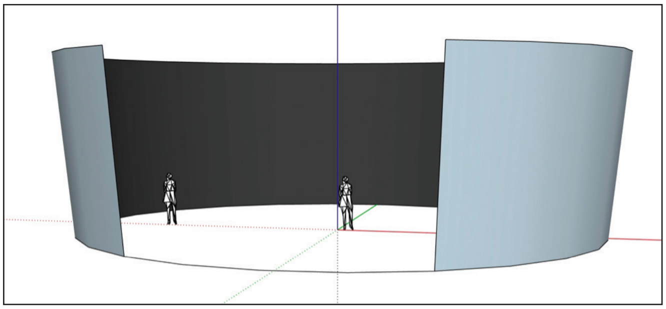


FIGURE 7. A depiction of a volumetric stage shaped in a conical frustum. (Imagecourtesy of Russ Berger.)

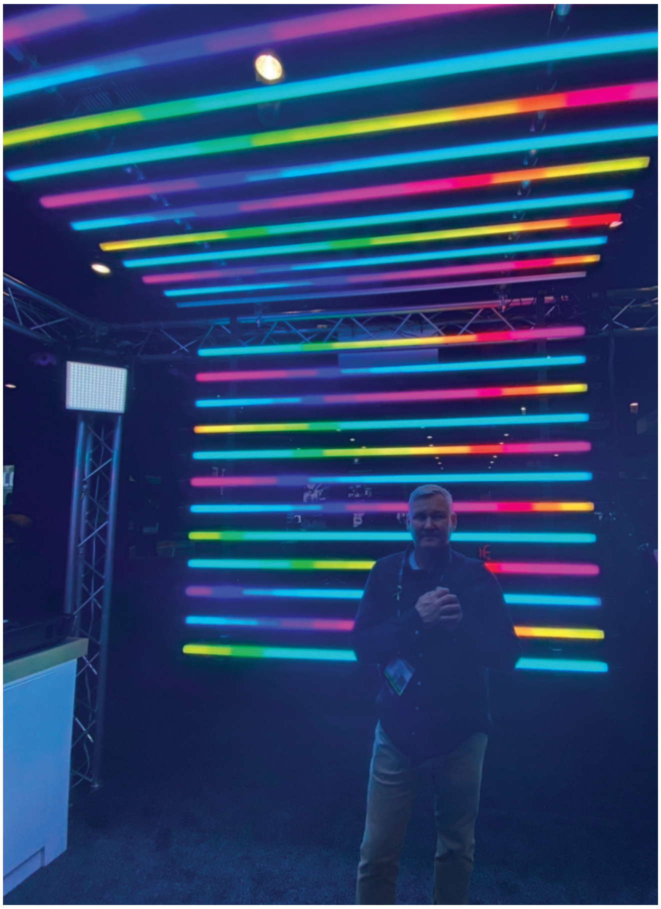


FIGURE 8. Image-based lighting example.

LED Ceiling: LED ceilings can range from complete coverage of a 360^o stage to partial coverage of whatever shaped stage, to portable ceilings to no ceiling at all. Ceilings facilitate a lot of sound reflection. Sound bounces up to the ceiling and back from side walls and the floor. Less ceiling equates to less reflection. The ceiling provides several options for sound reflection mitigation.

a) *Opening and Angling:* Russ Berger suggested making the ceiling panels movable, shaping the ceiling. For example, rows of panels could be opened like a Venetian blind (**Fig. 9**), with panel surfaces angled in rows toward the opening of the stage. Pitching a panel's surface toward the camera may direct a better quality of light and imagery than when panels are positioned parallel to the floor. Allowing the panel rows to open and close provides artistic flexibility. For car scenes, a flat, closed ceiling, without visible breaks in the reflective surface of the car, is preferred. For a close-up dialog scene when ceiling reflections aren't as visible, the panels could be more open. Some stages can partially or completely

1545-0279/24©2024SMPTE

retract their LED ceilings from the volume. Note that because cameras rarely capture virtual environment imagery directly from the ceiling, most ceilings are installed with less expensive LED panels with greater distance between diodes. The distance between diodes is described in millimeters and is termed a panel's 'pitch.' The greater the pitch value (distance between diodes), the lower the image resolution. Pitch in this context is not related to the angle of the panel's surface, as described earlier in the "LED Walls" section.

b. *LED Screen:* Some stages have their LED ceiling panels made with every other LED pixel removed. This creates a screen effect through which sound can exit and external lighting can pass through. LED screens offer a more limited reflective lighting effect as compared with LED panels that are fully populated with emitters. This solution works better on tall ceilings rather than short ones. On lower ceilings, such as car processing, the lines of missing LEDs can be noticeable in ceiling reflections.

c. *Image-based LED Lighting Alternatives:* As described above in LED Walls, when ceiling image detail is less of a requirement, these specialized LED lights provide better color quality light than LED panels, and in some cases (**Fig. 8**), with minimal acoustic reflection.

Equipment Room & Keyboard, Video, Mouse (KVM): The acoustic challenges of a volumetric stage include both reflection and unwanted noise intrusion. As described, volumetric stages are acoustically reflective, especially circular stages. Circular stages have an opening or two. Like parabolic microphones, sound that enters a circular stage from an opening can be focused and magnified. Actions occurring in front of a volumetric stage opening warrant concern that their sounds will be picked up, reflected, magnified, and recorded. Relative to traditional media production processes, OSVP requires additional equipment and personnel. Housing all processing equipment in an equipment room constructed to data industry standards not only keeps noise out of the stage, it's also better for equipment performance. KVM switches allow operators to interface their computers in the equipment room remotely. Additionally, unmanned servers or processors supporting the volume stage and any traditional supporting production equipment (DIT, transfer stations, etc.) can be housed in the equipment room, isolating heat and noise from the stage. Other media converters or extenders are available.

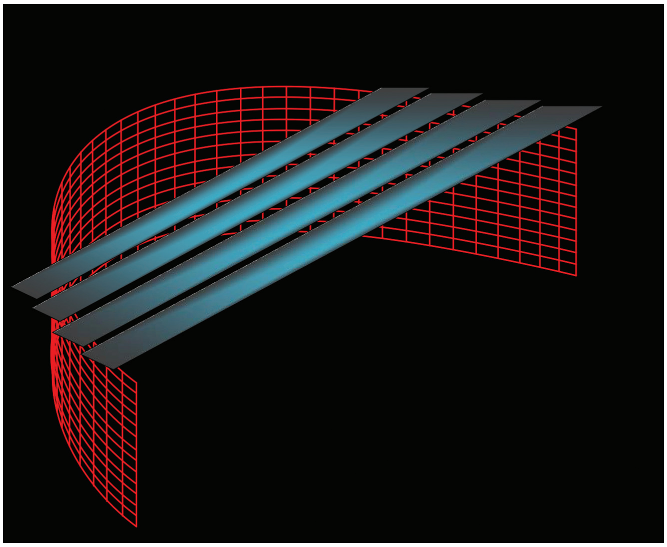


FIGURE 9. Example of LED volume ceiling panels configured in a Venetian blind pattern.

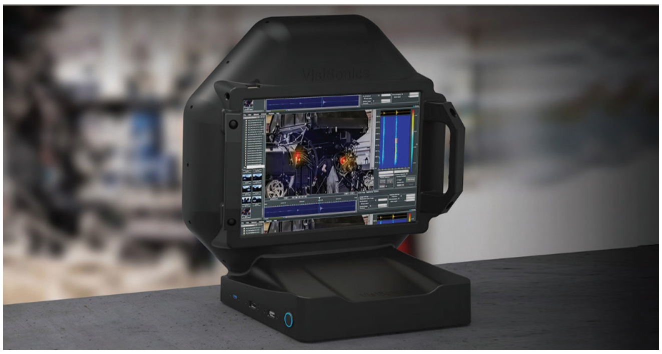


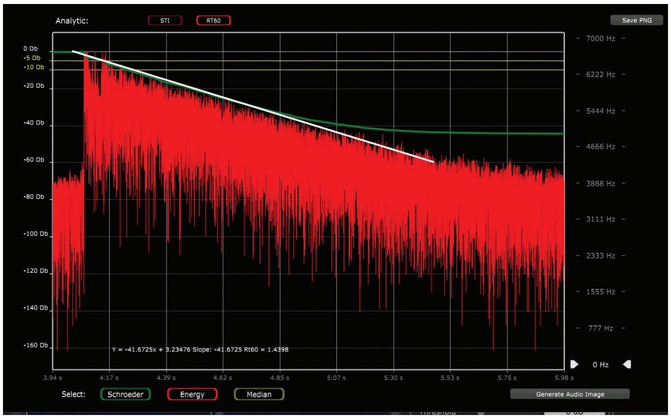
FIGURE 10. Example of a portable acoustical camera.

Visualization (Previs/Techvis): The visualization process provides the most significant opportunity to mitigate unwanted sound reflection and noise. Visualization occurs in preproduction, before production commences, the least expensive segment of production (production being the most expensive segment followed by postproduction). In visualization, each department can see the obstacles facing the other disciplines and can collaboratively work to help each achieve its creative goals, serving the overall creative vision. Often missing from the OSVP visualization process is production sound. The absence of sound within visualization demonstrates the continued hierarchical model of visual superiority.

Cooperation and Collaboration: a benefit of performing a robust visualization in the virtual world is the capability of the various creative departments to choreograph their stage setups. It also allows the opportunity to collaboratively resolve potential issues among the creative departments, maximizing overall creativity. As one production sound mixer/recordist said, "If they're cutting wood, it's too late [to make changes]."

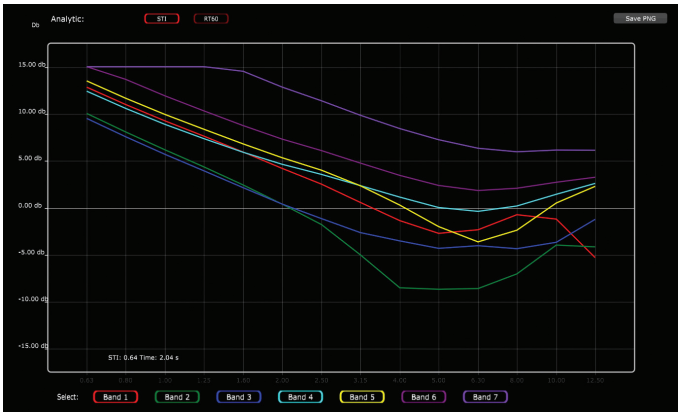
Greater Awareness of Sound Mitigation Efforts: In addition to sound reflection mitigation being practiced virtually, the visualization process creates awareness among the other departments of the challenges facing sound and the potential deployment of mitigation tools.

Production: Visualization complete, it's 'showtime.' What steps can be taken on set to mitigate sound reflection and



Y = -41.6725x + 3.23476 Slope: -41.6725 Rt60 = 1.4398

FIGURE 11. Acoustical camera RT60 calculation (1.44 sec) of balloon pop sonic instance recorded in an empty LED volumetric stage with DIY acoustical mitigation elements.



STI: 0.64 Time: 2.04 s

FIGURE 12. Acoustical camera STI calculation (0.64) of balloon pop sonic instance recorded in an empty LED volumetric stage with DIY acoustical mitigation elements.

noise in a virtual production volume stage that weren't addressed in visualization?

Personnel Removal: LED volume stages are magical. Everyone wants to be in or near it to see the technical wizardry happen as it happens. But everyone is a potential noise generator in an acoustically sensitive environment. Minor movements that may have occurred acoustically unnoticed on a traditional set now risk being captured within the curved reflector that is a volume stage, magnified by the reverberation, and thus recorded.

a) Clear the Stage: During shooting, minimize how many people remain inside the volume: hair & makeup, costume, set & props, grips & gaffers, etc. Resist the temptation to install a small video village in the volume. Convenient though it may be, whispers, tapping, the shifting and creak of a director's chair, the drop of an object, these sounds all get picked up and recorded.

b. Clear or Cover Stage Opening: Keeping the stage opening clear of non-essential personnel (those skills not requiring close proximity to the camera) and equipment will help to abate unwanted sound intrusion within a volume. A circular volumetric stage behaves similarly to a sensitive microphone. Any noise generated in front of it will be captured

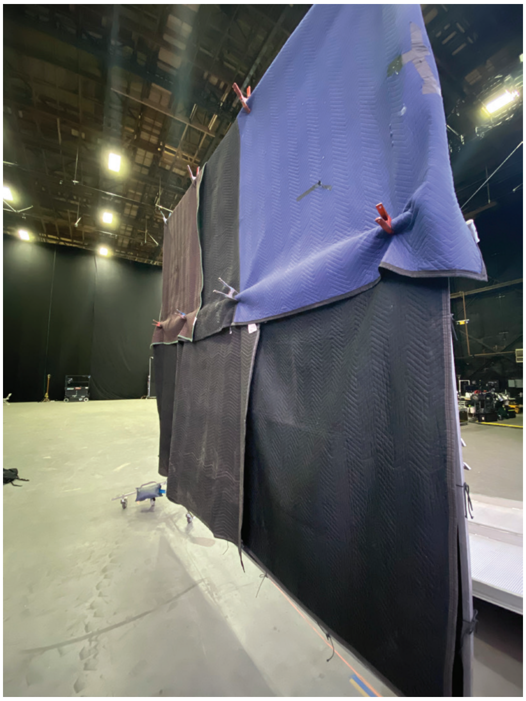


FIGURE 13. Simplest DIY acoustical shade, constructed of furniture blankets, hung over a frame and held up by two (2) C-stands. The one pictured is 12 ft by 12 ft.

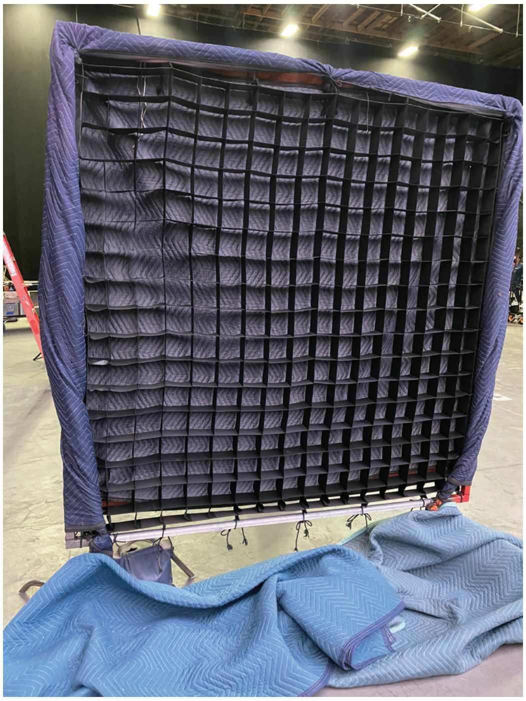


FIGURE 14. World's first DIY Echo Shade. A layer of duvetyne might improve the performance of this highly effective reverberation mitigation screen.

1545-0279/24©2024SMPTE

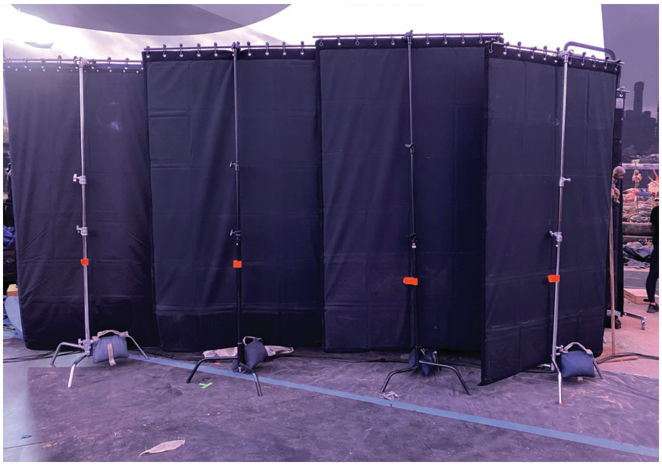


FIGURE 15. Zero-reflection (ZR) screens from Delta H. Designs, Inc. configured in a large LED volume stage, along with standard floor mats and furniture blankets.

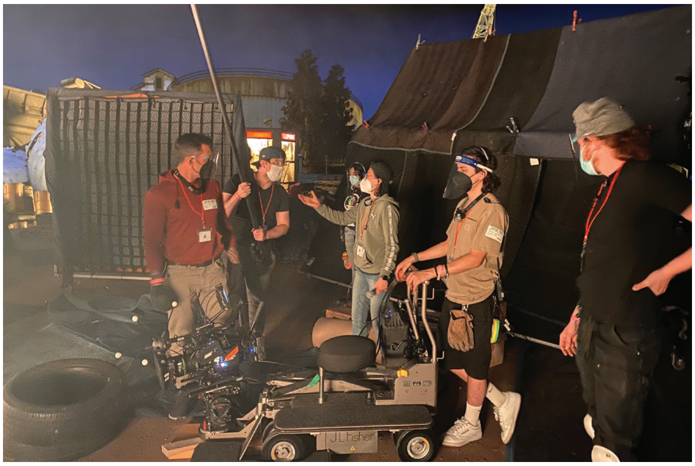


FIGURE 16. Two different types of DIY solutions, simple and echo shade, placed in an active set behind camera to capture sound originating from the actors and shield the actors from sound reflecting at them from the walls and outside the stage opening.

and potentially amplified. Additional OSVP personnel and equipment, including the virtual art department (VAD), sometimes referred to as the "Brain Bar" need not always be seated in, or in front of, the stage. With KVM, their workstations can be operated anywhere within the larger stage, away from the stage's opening. Covering the opening with a sonic barrier may be a less practical option. Boxing in or dividing noise generators within some sort of acoustic barriers may be a more cost-effective alternative.

Microphone Placement: With current non-array, non-intelligent microphone technology, there is little difference between how a production sound mixer/recordist would likely want to place his/her/their microphones as compared with traditional stage and location recording, an isolated lavalier microphone on each actor and a boom microphone as close as practical, pointed at the actor.

Mitigation Tools and Practices

a. *Pre-shoot Sound Sweep:* Locate and address undesirable noise generators before crews arrive and shooting begins. For those interested in high-tech solutions, an acoustic camera sees issues in realtime and can instantly locate origination points of noise (**Fig. 10**).

b. *Interdepartmental Communication:* Following an inclusive visualization process described earlier, clear, constant, collaborative on-set communication between the camera & lighting, the assistant directors, and sound is the next most significant sound reflection and noise mitigation practice to support.

c. *Mitigation Gear and Placement: Standard Sound Mitigation:* Production sound mixer/recordists already deploy materials to mitigate sound reflection: furniture blankets (furni-pads) and screens made of furniture blankets (described later under "Sound Screens/Shades"), carpeted rubber floor mats, foam insulation, taping shoe soles, and more. To mitigate sound reflection in a volume stage, more of these same materials need to be utilized to cover more area than in traditional sound stage set recordings.

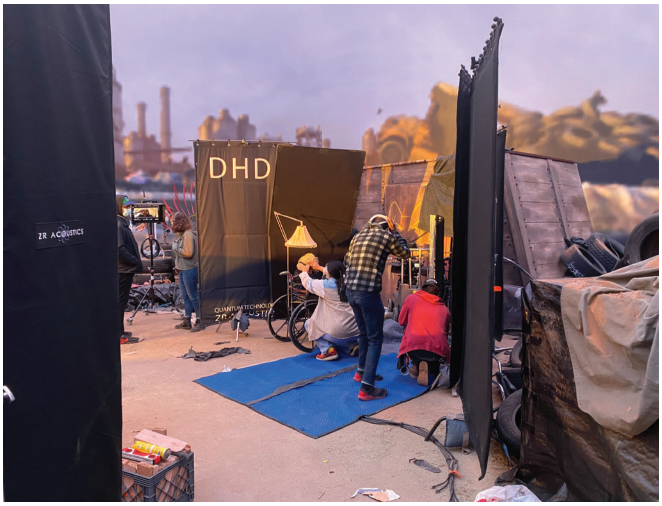


FIGURE 17. Zero-reflection (ZR) screens placed in an active set behind camera to capture sound originating from the actors and shield the actors from sound reflecting at them from the walls and outside the stage opening.

Sound Screens/Shades: In addition to the standard sound abatement materials, acoustical mitigation elements specific to high-reflection environments can be set up to prevent sound from reaching the wall and/or prevent sound reflected off the wall from reaching the actor. In a balloon pop test, a do-it-yourself echo shade and the Delta H. Designs' ZR screens performed similarly, cutting reverberance in half, from an RT60 of 3.2 sec in an empty large volume stage to between 1.4 sec and 1.7 sec in the

1545-0279/24©2024SMPTE

same empty stage. The speech transmission index also improved, from 0.51 to 0.64-0.72 (**Figs. 11 and 12**). These screen elements are typically held up on sandbagged C-stands, with horizontal arms or a panel frame holding them open. Examples of 3 acoustical screen types are described next: two do-it-yourself (DIY) and one manufactured.

Simplest (Do-it-Yourself (DIY)): Sound shades of duvetyne or furniture blankets are standard sound abatement equipment. The grip department could hang duvetyne or furniture pads from a panel frame held up by two C-stands (**Fig. 13**) or from a single C-stand with a horizontal arm (holding material about 4 ft wide by 8 ft tall).

Echo Shade (DIY): A novel, slightly more complex but more effective DIY solution is an echo shade (ES). Assembled similarly to the previous DIY example using a panel frame but with an added feature. In lighting, polyester honeycomb grid diffusers are commonly used and available on set. Taking the above DIY hanging blanket solution described earlier using a panel frame and two C-stands, attach in front of it one of the lighting's polyester honeycomb grid diffusers (**Fig. 14**).

Zero-Reflection (ZR) Screen: Available for purchase or rent from Delta H Designs Inc. is a zero-reflection (ZR) screen. These foldable screens are unfolded to about 4 ft wide by 8 ft tall and held up off the ground by a sand-bagged C-stand stretched out across and affixed to a horizontal steel, arm bar are said to offer better performance (**Fig. 15**).

Screen/Shader Placement: Regardless of which screens or combinations of screens, positioning is the same. Screens should be placed as close to a speaking actor as camera framing (picture visible to the camera sensor) will allow so that no images of the screens are captured or recorded. Screens should be faced toward the actor's mouth when speaking. For example, in an experimental OSVP project, during the first week, we deployed a 6-ft by 6-ft ES screen close to actors and a 12-ft x 12-ft dark furni-pad panel frame screen nearby (**Fig. 16**). The second and third weeks, the 6 ft x 6 ft echo shade (ES) was replaced by nine ZR screens with the 12 ft x 12 ft screen remaining in action (**Fig. 17**). Any screens that could not be placed near the speaking actor, 'catching' their voice, were either stood up a little further away, but still as close as possible to prevent sound from reaching the wall or ceiling, or in the opposite direction, facing whichever wall would most likely reflect sound toward the actor (**Fig. 18**).

Sound Mitigation Labor: A sound mitigation laborer added to the production crew provides dedicated effort toward tackling the acoustical challenges facing OSVP, allowing production sound recordists/mixers time to dedicate toward their art and other responsibilities instead of time-consuming labor tasks.

High-Resolution Recording: Post-production processing always degrades the quality of sound. The higher the sample rate and bit depth of the production recordings, the more margin (bandwidth depth) post-sound editors have for guarding performance after processing the dialog tracks with such tools as reverberation removal software. Currently, 48 kHz, 24-bit is the common rate for production sound mixers/recordists. For OSVP, 96 kHz, 24-bit is a recommended minimum, but by no means an absolute rule.

Post-Production

Post-Production Sound Applications: Currently, while several echo, reflection, and de-reverberation applications and plug-ins exist, iZotopeRx seems popular and effective in removing unwanted reverberation. As is the case with

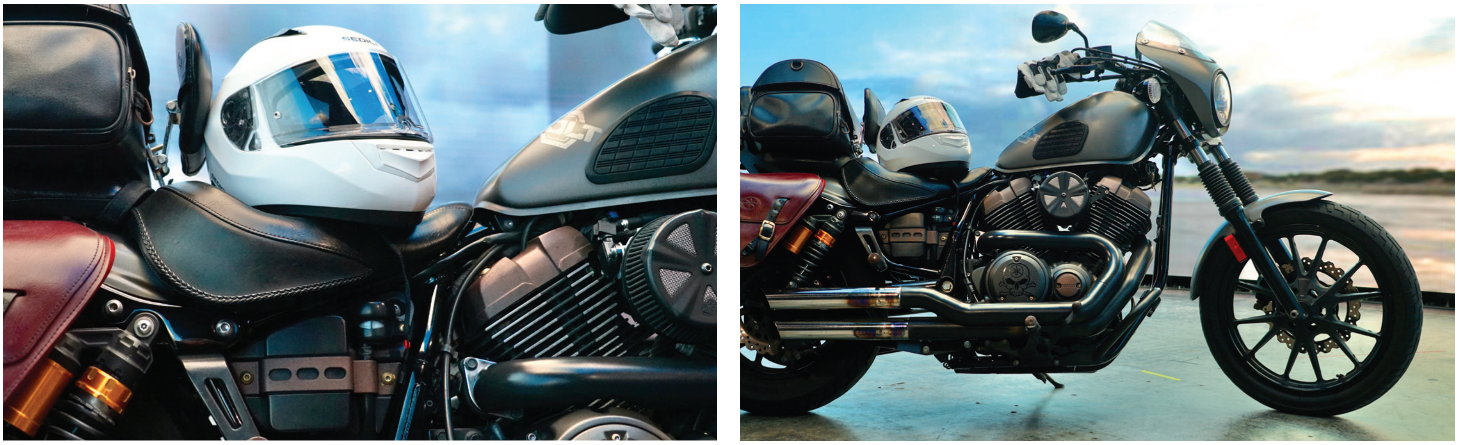


FIGURE 18. Environmental reflections of LED imagery on actors, costumes, props, and set pieces are one of the sought-after advantages associated with shooting in an LED volume in an LED volume stage.

Glossary of Acronyms	
ADR	Automated Dialog Replacement
AI	Artificial Intelligence
BGs	Background Sounds
DIT	Digital Intermediate Technician
DIY	Do It Yourself
DP	Director of Photography
ES	Echo shade
ETC	Entertainment Technology Center
IBL	Image-based Light
ICVFX	In-camera Visual Effects
KVM Switch	Keyboard, Video (monitor), Mouse
LED	Light Emitting Diode
OSVP	On-set Virtual Production
RGB	Red, Green, Blue
RT60	Reverberation Time 60 decibels (ISO 3382, the duration required for the sound energy in a room to decrease by 60 dB after the source emission has stopped.)
STI	Speech Transmission Index
USC	University of Southern California
VAD	Virtual Art Department
VFX	Visual Effects
VP	Virtual Production
ZADR	Zero Automated Dialog Replacement
ZR	Zero-Reflection

any creative technology, the artist plays the largest role in the effectiveness of a tool or discipline. De-reverberation is no exception. The ability to repair audio juxtaposes several variables: the quality of audio recorded (the higher, the greater the manipulation latitude), the degree of reverberation generated within a given environment (is it within reasonable range of repair), and the value the director and producers place on a performance (the tolerance level, original performance vs ADR).

Background Sounds (BGs): A common creative technique used among sound supervisors, editors, and re-recording mixers is to bury the last bit of unwanted levels of reverberation under believable background sounds, thereby masking over the reverberation with ambient (background) sounds or music.

Re-recording Mixers: The final and most costly per-hour part of the sound mastering process should be a production's last resort to address the reverberation recorded in dialog tracks. Apart from mixing in backgrounds to bury reverberation in dialog tracks, re-recording mixers, just like sound editors, can use applications like iZotope's Rx to lessen the reverberation effect. Again, not all re-recording mixers are equally talented regarding any specific skill, including dereverberation removal. But having re-recording mixers remove reverberation in a mix stage can, and is, done.

Conclusion

The quality of dialog sound often determines how audiences perceive a performance. OSVP stage designers, builders, and facility owners have introduced highly reverberant production environments within which the production expects the production sound mixer/recordist to work. The steps required to successfully mitigate sound reflection and unwanted noise in an LED virtual production stage are the production's responsibility. Through preparation, collaboration, detailed visualization, additional equipment and personnel, and supportive relationships among all the on-set creative professionals, productions can succeed in capturing original performance dialog that carries over to the final sound mix, significantly minimizing ADR usage. In the future, acoustic-friendly stage construction combined with sound mitigation specialists, acoustical education, and improvements across all production departments will deliver to audiences the vocal performances directors, actors, and sound professionals strive to deliver, providing better results more quickly at costs producers can accept, offsetting ADR costs.

To hear and see examples of sound in an LED volumetric stage, visit www.finalsample.com.

Acknowledgments

This preliminary study was made possible by the generous support and collaboration of many brilliant individuals, groups, manufacturers, and virtual production studios. Through it all, my wife, Leslie Dennis Rigney, remained my greatest support.

- *Professionals:* Sean Tajkowski, Russ Berger, Kami Asgar, Chris Robleto-Harvey, Michael Koffy, Mark Ulano, Patrushkha Mierzwa, Shawn Holden, Paul Debevec, and Nickolas Rigney and Bryce Rigney
- *ETC/FATHEAD Production:* Erik Weaver, Ben Baker, Greg Ciaccio, c.Craig Patterson, Brandyn Johnson, Pablo Mena, Miranda Friel, Maddi Mays, Mitchell Colley, Anthony Gaitros, Becca Han, Letia Solomon, Alexa Villareal, Tom Thudiyanplackal, and the Camera, Lighting, Grips and Gaffers crew.
- *Manufacturers:* VisiSonics, Adam O'Donovan; Dotterel Technologies, Shaun Edlin; Delta H. Designs, Inc, Hanson Hsu and Sherri Holmes; Planar, Richie Argue; Quasar Science, Tim Kang; ARRI; Epic Games; Nvidia; Intel; and WACOM
- *Stages:* XR Stage, Orbital Virtual Studios, Nant Studios, Amazon Studios, and Vū Stages

About the Author

Eric Rigney, lead instructor and curriculum developer at Mo-Sys Virtual Production Academy, is an active participant in the University of Southern California's Entertainment Technology Center's virtual production initiatives, SMPTE's RIS OSVP educational development, and the Hollywood Professional Association.

Presented at the SMPTE 2022 Media Technology Summit, 24–27 October 2022.

DOI: 10.5594/JMI.2024/WXMY3468
Date of publication: 24 January 2024

Standards Technology Committee Meetings

On a quarterly basis, the Standards Community convenes for week-long TC Meetings. During these sessions, participants provide updates on progress and collaborate on advancing standards work.

4-7 March 2024
ONLINE

3-5 June 2024
OTTAWA, CA

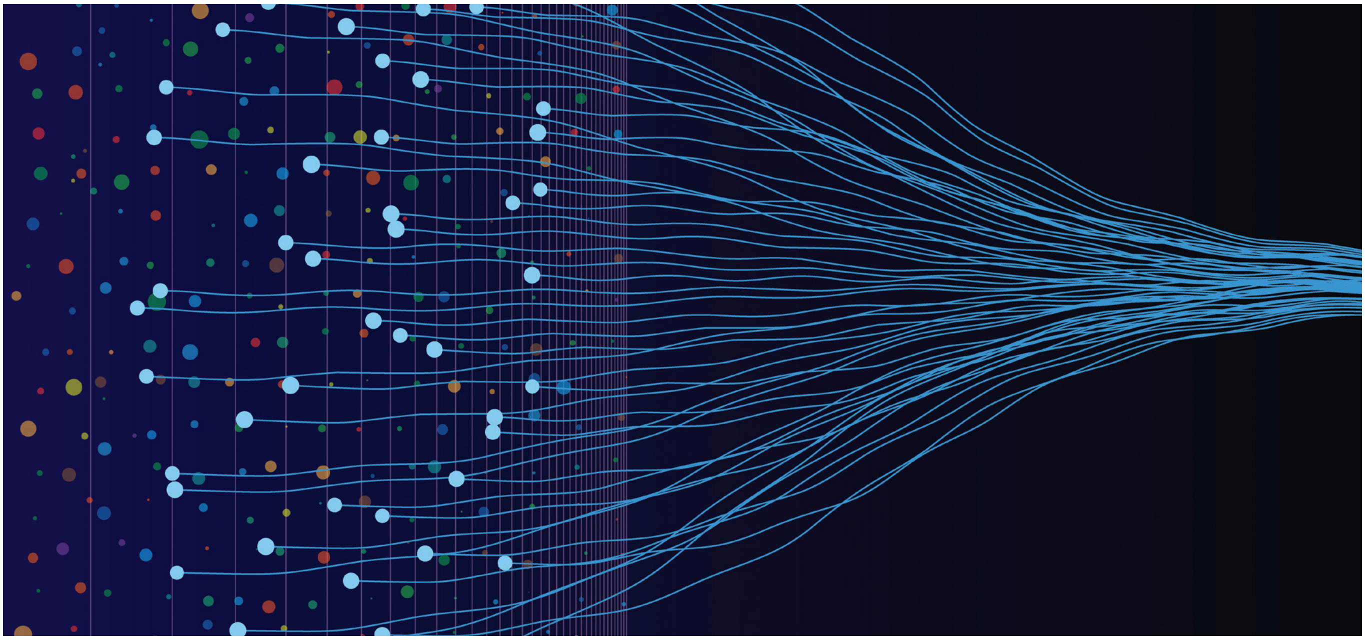
18-20 Sept. 2024
GENEVA, CH

Interested in hosting a TC Meeting?

SMPTE

OFF-TOPIC: OPINION

KEYWORDS CONTENT CURATION // GENERATIVE AI // INDUSTRY ARCHITECTURE // HYPER CONTENT PERSONALIZATION



The Future Vision of Content Personalization Through Trusted Curation

By John Footen, Jason Williamson, Blake White, Jesse Pitt, and Garrett Coley

Highly personalized content, which resonates emotionally in many ways, delivers a more engaging and memorable entertainment experience. Once this capability fully matures, producers could generate different versions of programs for viewers from the raw and finished content essence and metadata.

1545-0279/24©2024SMPTE

Abstract
Future Content Personalization techniques will likely flip the power dynamic between centralized content producers/distributors and networked media consumers. Critical to this power shift will be personalized control of all consumer data and subsequent viewer preference matching that will impact how content is produced, how data is managed, and how advertising is delivered. This paper envisions a new conceptual and viewing-context-aware AI-enabled entity in the media ecosystem—the "Curator Agent." We explore the required innovations for this future—fully trusted technology to enable hyper-personalized and context-aware content to become a reality.

Today, personal preferences and viewing history do not follow when a viewer switches between streaming providers. Each platform algorithm deduces viewer preferences from behaviors exhibited only on that platform. This is done over time while content consumers are presented favorable and unfavorable content options.

What if comprehensive viewer preferences followed the consumer across commercial platforms? Digitally produced and distributed content opens the door for consumers to take direct advantage of the power of the cloud to enhance their entertainment experience. Use of AI-enabled preference mapping will make content experiences compelling. Edge networks and consumer devices will be expected to gather and securely maintain highly contextualized profile data to deliver uniquely personalized content regardless of the service provider.

In this scenario, a secure "Curator Agent," owned and controlled by an individual, could allow consumers to control and selectively share dynamic content preferences and associated personal insights across all service providers. This would put the viewer in the driver's seat, curating their content experiences and perhaps someday personalize how one experiences a particular title or show.

Highly personalized content, which resonates emotionally in many ways, delivers a more engaging and memorable entertainment experience. Once this capability fully matures, producers could generate different versions of programs for viewers from the raw and finished content essence and metadata. The graphics could be personalized, but so can objects on the screen, characters, voices, dialog, volume, and the advertisements served to a consumer or audience (in screen or interstitial).

Of course, this level of personalization requires data, context, and insights describing a consumer; not only who they are, but when they are watching, what they experienced that day, or any other personal information they are willing to make available to enhance their experience. This is where the curator agent enters the equation.

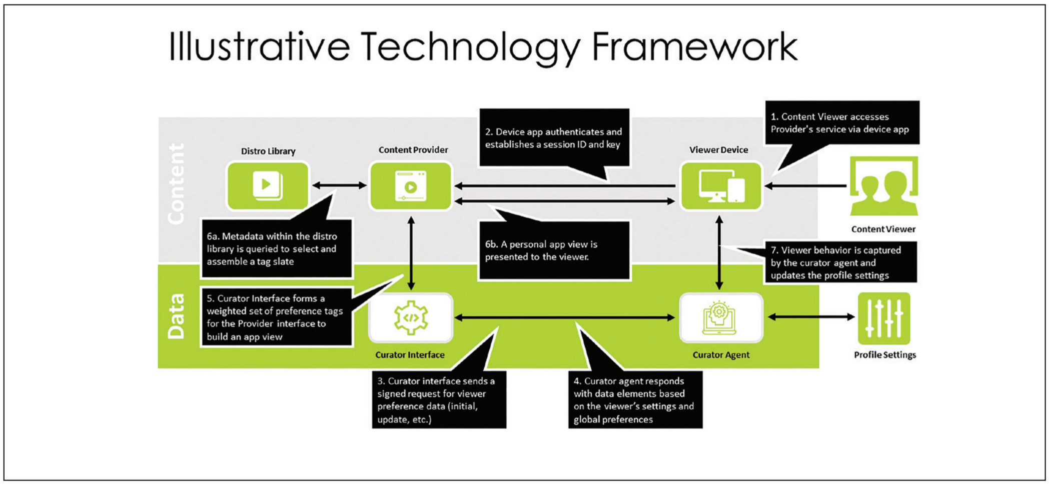
Costs and the lack of an existing, trusted, third-party entity, make it unclear when this entity will emerge. Still, as costs decrease, an examination of the potential for Curator Agents is valuable.

Trust Creates a Win-Win Content Ecosystem

The industry must adopt a new trust-based ecosystem alongside beneficial business models between producers and consumers to encourage the forward evolution of hyper-personalized content. Centralized aggregation and control of consumer data by content producers, distributors, and advertisers is the heart of the problem. Our research[1] reveals that 47% of consumers say they trust their online services to protect their data, signaling a desire among some for more privacy:

1. Consumers will take steps to protect their privacy when they have easily accessible options. This sentiment has been revealed in Deloitte's research since 2018.
2. Many consumers would consider paying for more control over their privacy. According to Deloitte's Digital Media Trends[2] survey, 57% of consumers said they would be willing to pay for the ability to view and potentially delete personal data that companies collect, and 45% would be willing to pay to access a social media platform if it agreed not to collect personal data.
3. A Digital Media Trends Survey also concluded that advertisers are still placing most ads with little apparent relevance to the consumer at the moment of viewing. It can be argued that this is due, in part, to a lack of up-to-the-minute information of the viewer's recent experiences and viewing context. Current approaches to demographics and other viewer-relat-

1545-0279/24©2024SMPTE



Illustrative Technology Framework

1. Content Viewer accesses Provider's service via device app

2. Device app authenticates and establishes a session ID and key

3. Curator interface sends a signed request for viewer preference data (initial, update, etc.)

4. Curator agent responds with data elements based on the viewer's settings and global preferences

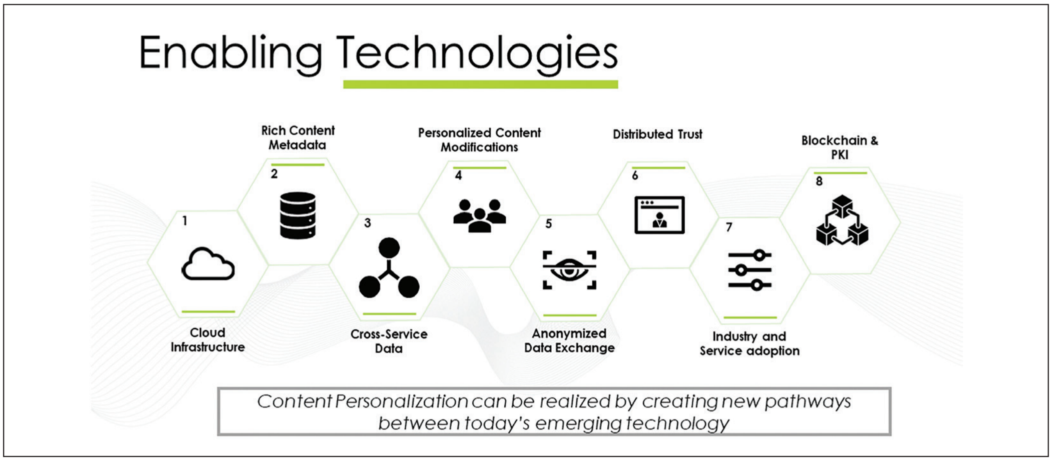
5. Curator Interface forms a weighted set of preference tags for the Provider interface to build an app view

6a. Metadata within the distro library is queried to select and assemble a tag slate

6b. A personal app view is presented to the viewer.

7. Viewer behavior is captured by the curator agent and updates the profile settings

FIGURE 1. Illustrative technology framework.



Enabling Technologies

Content Personalization can be realized by creating new pathways between today's emerging technology

FIGURE 2. Enabling technologies

ed data can be either too broad or stale to provide ad servers on what would create the biggest impact.

Consumers appear to be somewhat less concerned about sharing data that fuels the revenue engines of many tech giants and advertisers—such as browsing history, purchasing data, and social media activity. Consumers are willing to share an astonishing amount of data, including personal details, if companies are willing to deliver value in exchange. Companies can make consumers more comfortable about what they share by giving users control over their data.

Our research suggests that media companies focused on consumer autonomy, choice, and a privacy-first approach may gain a competitive advantage. Why not let consumers update their own profiles and share their changing aspirations and preferences in real-time, while at the same time, granting service providers limited, perishable consent to use the data consumers are comfortable sharing? This would make consumers partners in an exchange of value, rather than a commodity—an exchange that could be more valuable for businesses. Of course, the viewer-controlled personal preference data can be valuable to, and monetizable by, the viewers themselves as well.

A New Industry Approach

Creating a future in which content is so context-sensitive and specific to the viewer's preferences, identity, and desires while respecting consumer data privacy, will require developing innovative technologies and new business approaches in the industry (**Fig. 1**). Central among these technologies will be a new entity in the media ecosystem we call a curator agent—a decentralized, viewer-controlled, and trusted data vault that serves as an intermediary between a consumer and media service provider. The personal data required to deliver a truly personalized experience will no longer be collected and maintained by service providers. Rather, content consumers will own and control the access of their collected data along with the ability to share their profile across providers of their choice. Responsibility will, therefore, shift to the content consumers, significantly lowering content provider legal and regulatory risks and overhead.

The viewer controls the ability to share varying aspects of the data collected with content producers. A viewer may choose to share two subsets of information with one provider (e.g., location and basic demographics) and all data with another. Content providers are likely to develop unique benefits to encourage data sharing from the viewers. However, the agent can filter any data prior to sharing it with content service providers since the viewer manages this data access.

Viewers will also control their engagement experiences with transparency of the collected data about their engagement and preferences. Curator agents may collect data across many dimensions of viewers' lives, as permitted. This agent would allow the content to be updated in realtime. For example, if someone's heart rate sped up, then the content could focus on that part of the show. If there is a sporting event and the viewer keeps repeating a player's name, they could be offered more camera views of that player in real-time. This type of examination of the viewing circumstances (the context) would typically be seen as intrusive, but in this case is carried out with the assistance of a trusted agent who will only share this data in the manner and to whom the viewer chooses.

Enabling Technologies are Rapidly Evolving

Creating these experiences requires continuing evolution in cost and capability in several underlying technologies, including the following (**Fig. 2**):

1. **Cloud Computing:** Cloud and its service-based elastic infrastructure extending to the edge is key to realizing this vision. Powerful edge and consumer devices will play a larger role in helping to enable the content modification process, implementing just-in-time experience enhancement and content assembly for presentation to the consumer. Computation efficiency, storage density, bandwidth efficiency, and certain forms of compression will enable more sophisticated and engaging features as these technologies mature.
2. **Rich Content Metadata:** Media capture and intelligent metadata tagging are the foundation to create multiple views and personalized experiences for audiences. Envision today's multi-service experience as an ocean of content. Systems answer highly personalized requests by selecting from an exceptionally large volume of content. A future state exists where assembling smaller sets of content offerings enables providers to present a more precise and relevant slate of content.
3. **Cross-service Data:** A unique outcome of a curator agent is that recommendations are made better with central access to cross-service libraries and usage data. As the volume of behavior and preference data grows, content is directed and aligned to a viewer's preferences in an increasingly accurate and sophisticated manner.
4. **Personalized Content Modifications:** Alongside recommendations, generative and other types of AI could stitch content together based on interest profiles to create a more personalized narrative. For example, when creatively appropriate, B-roll footage metadata and consumer preference profiles could be leveraged to assemble different storylines or avoid objectionable content.
5. **Automatic Profile Updates:** Bidirectional data exchange is required to establish a secure feedback loop to determine whether viewers acted upon the recommended selection. Data exchange might be modeled after existing data workflows, but ultimately the maintenance of this data toward an accurate and fresh view of a consumer's preferences is key to optimizing the experience.
6. **Distributed Trust:** A key requirement is a multiple signature approach in which both the viewer offers their data for personalization, and the organization interested in receiving their anonymized data can interoperate. This data could be anonymized or obfuscated to form that personalization element without exposing the raw data that went into creating the personalization element.
7. **Industry and Service Adoption:** While control of, and access to, dispersed networks of consumer/viewers' specific profile data will be increasingly valuable, it will be difficult to get away from the current highly concentrated or centralized mode of preference aggregation algorithms and scalable distribution platforms with new, reasonable business models. This will require interoperability standards that are best addressed by standards development organizations in collaboration with service providers.
8. **Blockchain:** Blockchain technologies are designed to be decentralized, anonymous, and enforce trust mechanisms between untrusted parties. However, many blockchain-based projects became centralized to ease user adoption and maintain founder control, thus sidestepping the core benefits of this technology. While blockchain was examined as a potential tool, we do not believe blockchain technology is initially required for our vision. A secure database and public key infrastructures can fulfill many core security and privacy needs on faster, commoditized, and more broadly supported platforms.

A Curator Agent Allows New Opportunities

Like every other disruptive technology innovation, this approach will create new opportunities for every player in the media ecosystem. Even the existing business models will be enhanced by creating greater engagement with content and more relevance for advertising. The power of content being aware of viewing context has not been widely explored to date, but is almost certain to be more effective.

Broadcasters and advertisers will have agents that interact with the curator agent to satisfy content interests and monetization opportunities. However, a viewer's agent will be designed to work on the viewer's behalf. The agent can search the ecosystem for content that will deliver high value based on what it knows about the viewer. Content producers will request temporary and appropriate access to the consumer-approved data. Content producers will use the shared data to determine the best show and customizations for the viewer. All data from consumers and creators will be analyzed to provide feedback to the content producers as well as advertisers.

A curator agent will interface with media companies' platforms to help them mix and remix fresh content, delivering different versions of a show based upon mood, level of interest in details, affinity for talent and influencers, the perspective of a favorite sports legend, and a wide range of emotional aspects of which your secured personal profile curator agent is aware.

The potential for influencers to sell or give away their preferences to their fans would create new business models for them. Fans would be able to mix those preferences with their own to help them see the world through the eyes of others. This might be extended to include influence from other entities that a viewer might have an affinity for, such as clubs, religious institutions, trusted civic entities, and more.

It could also spark the growth of more content players in a growing content ecosystem, such as companies developing and licensing digital content assets consumed across multiple platforms, windows, geographies, and audience segmentations. These smaller digital assets can include smaller/shorter segments, scenes, stories, 3D models, animation, sound effects, AI/ML-enabled behaviors, dynamic action algorithms, virtual scenes, virtual characters, AI actors, computer-generated audio/speech, and localized versions.

These assets also show the ability to reuse and create derivative works that can be monetized by creatives working outside of the traditional ecosystem. Emerging players might include social media publishers, independent creatives, stock footage houses and advertising agencies that desire to use content. New companies that specialize in specific types of realtime customizations could emerge.

All of this will almost certainly lead to new genres of content never seen before. The personalization ecosystem has the potential to generate the most compelling version of documentaries, dramas, sports recaps, news summaries, movies, and advertisements that match viewer interests—and this precise relevance will enhance engagement and monetizability by producers. The creative implications of these new capabilities are fascinating to consider and are reminiscent of the new creative fields that were enabled by the computer gaming revolution.

Conclusion

To accomplish this vision of content personalization, core enabling technologies must further evolve in capability and cost, and they are rapidly on a path to doing so. Nevertheless, we speculate that foundational technologies are sufficiently advanced to build a proof-of-technology for this concept, and that the next-step is to prototype the solution illustrated in **Fig. 1**.

Like other significant technology innovations, this will disrupt how content is conceived, produced, distributed, presented, consumed, and monetized. This will be a further disruption of traditional media. As we currently experience it, the industry will evolve to be targeted at the device and even viewing-specific context level and make content and advertising more valuable. The disruption will also create opportunities for a new generation of content service companies and the associated technology providers.

Will the media industry embrace and enable this inevitable change? Or will it resist by holding on to historical legacy "push" models of mass audience distribution? If the history of modern technology innovation teaches us anything, it is that self-disruption is far less painful than externally induced disruption by new unforeseen entrants.

References

1. Jeff Loucks. *First control, then consent. Consumers will share personal data if they're in charge of how it's used*. Deloitte, Thinking Fast, September 6, 2018. [Online]. Available: https://www2.deloitte.com/us/en/pages/technology-media-and-telecommunications/articles/personal-data-privacy-regulations-and-consent-of-consumers.html
2. Kevin Westcott et al., *Digital media trends, 15th edition: Courting the consumer in a world of choice* , Deloitte Insights, April 16, 2021.

About the Authors

John Footen is a managing director at Deloitte's Media Solution Practice, where he lead the company's Media Solutions practice with over 25 years of experience in the Media & Entertainment (M&E) industries. A SMPTE Fellow, he is an industry expert and Emmy winner.

Jason Williamson is managing director at Deloitte Consulting, where he brings over 20 years of experience in the Media & Entertainment and Technology industries, specifically on digital content distribution, asset management, production workflows, broadcast, and operations strategy.

Blake White is specialist leader at Deloitte Consulting LLP. Leveraging extensive Silicon Valley product and Hollywood service competencies prior to joining Deloitte in 2020, White has more than 25 years of digital technology and business transformation experience in North America, Europe, Africa, and Asia.

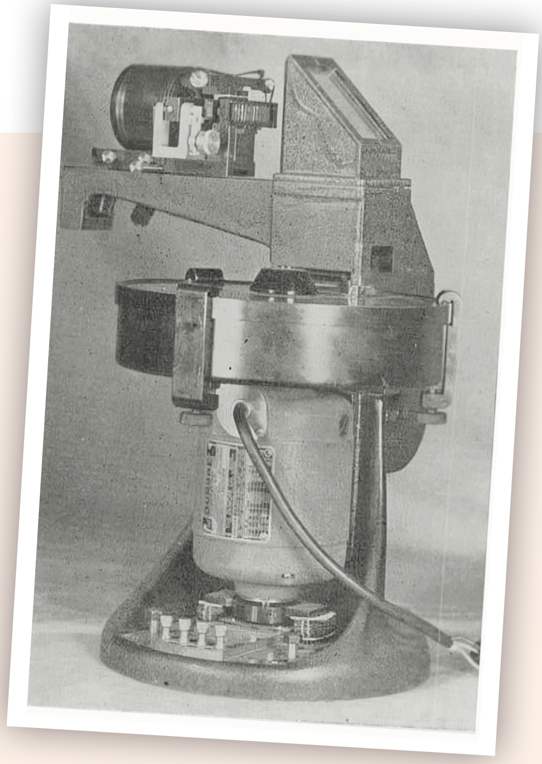
Jesse Pitt is specialist master at Deloitte Consulting. He has over 20 years of Live sports and special events accomplishments. He enjoys leading Media Solution clients through broadcast transformations, AI/ML enriched cloud enablement programs, across media supply chain services from lenses to living rooms.

Garrett Coley is an analyst at Deloitte Consulting, where he has consulted and led a variety of cloud implementation and product development engagements. Coley is dedicated to empowering users to take ownership of their data and leveraging it for curated consumption of goods.

Presented at the SMPTE 2022 Media Technology Summit, 24–27 October 2022.

DOI: 10.5594/JMI.2024/INNK6533
Date of publication: 24 January 2024

THE ORIGINAL **JENKINS PROJECTOR**, NOW IN THE UNITED STATES NATIONAL MUSEUM, CONSISTED OF AN OPTICAL SYSTEM WHICH FIRST SUCCESSFULLY PROJECTED MOTION PICTURES ON A SCREEN.



MICHAEL DOLAN

25 Years Ago in the Journal

The February 1999 *Journal* published in: “Progress Report—Television” by Robert P. Seidel:” Nineteen hundred and ninety-eight was a milestone year for the television broadcast and production community as it began the transition to DTV. The Consumer Electronics show kicked off the year with prototype receivers being displayed by almost every manufacturer. Temporary broadcast facilities were set up by ABC, CBS, and PBS to provide receiver manufacturers with actual HDTV broadcast signals. Many observers indicated that they were disappointed that the manufacturing community had not progressed further in DTV receiver development. However, the manufacturers assured their detractors they would have product available for the 1998 holiday season...No sooner was the Olympics over than everyone was focused on NAB 1998. One did not have to walk

In this column, we provide interesting historical briefs from the journal articles of days past. The purpose of this column is primarily entertainment, but we hope it will also stimulate your thinking and reflection on the Society’s history, how far we have come in the industry, and (sometimes) how some things never change. This column is sponsored by Television Broadcast Technology, Inc., since March 2001.

1545-0279/24©2024SMPTE

far in the exhibit area to realize that HDTV had transitioned from the wings to center stage...In early spring, many stations began tower modification and construction projects for DTV...During the summer, the ATTC completed their measurements on the "Evaluation of DTV taboo channel interference into NTSC under strong signal conditions...A... milestone occurred on October 29, 1998, when the Harris Corp. sponsored the John Glenn Space Shuttle Launch that was carried live coast-to-coast on the CBS high-definition network consisting of eight early adopter stations."

50 Years Ago in the Journal

The January 1974 *Journal* published in: "Colorimetric Standards in U.S. Color Television" by L. E. DeMarsh: "The objective of current standards efforts in the U.S. television industry is to improve both the uniformity and quality of color television as viewed on the home receiver. U.S. broadcast colorimetric standards are based on the original NTSC primaries. However, the phosphors used in current television display devices are significantly different from those specified by NTSC. These phosphor differences result in a colorimetric mismatch that may cause significant color distortions in the display. Receiver manufacturers in the U.S. have adjusted the color demodulation characteristics in their receivers to compensate for these phosphor changes. To resolve this problem in Europe, the European Broadcasting Union (EBU) has recently recommended that broadcast standards be changed to require matrixing for modern phosphors in all cameras...After making a study of how this basic problem affects U.S. television, it appears to be preferable to retain the NTSC standards and to insert a correction matrix in the display device that will compensate for the particular phosphers in use.

75 Years Ago in the Journal

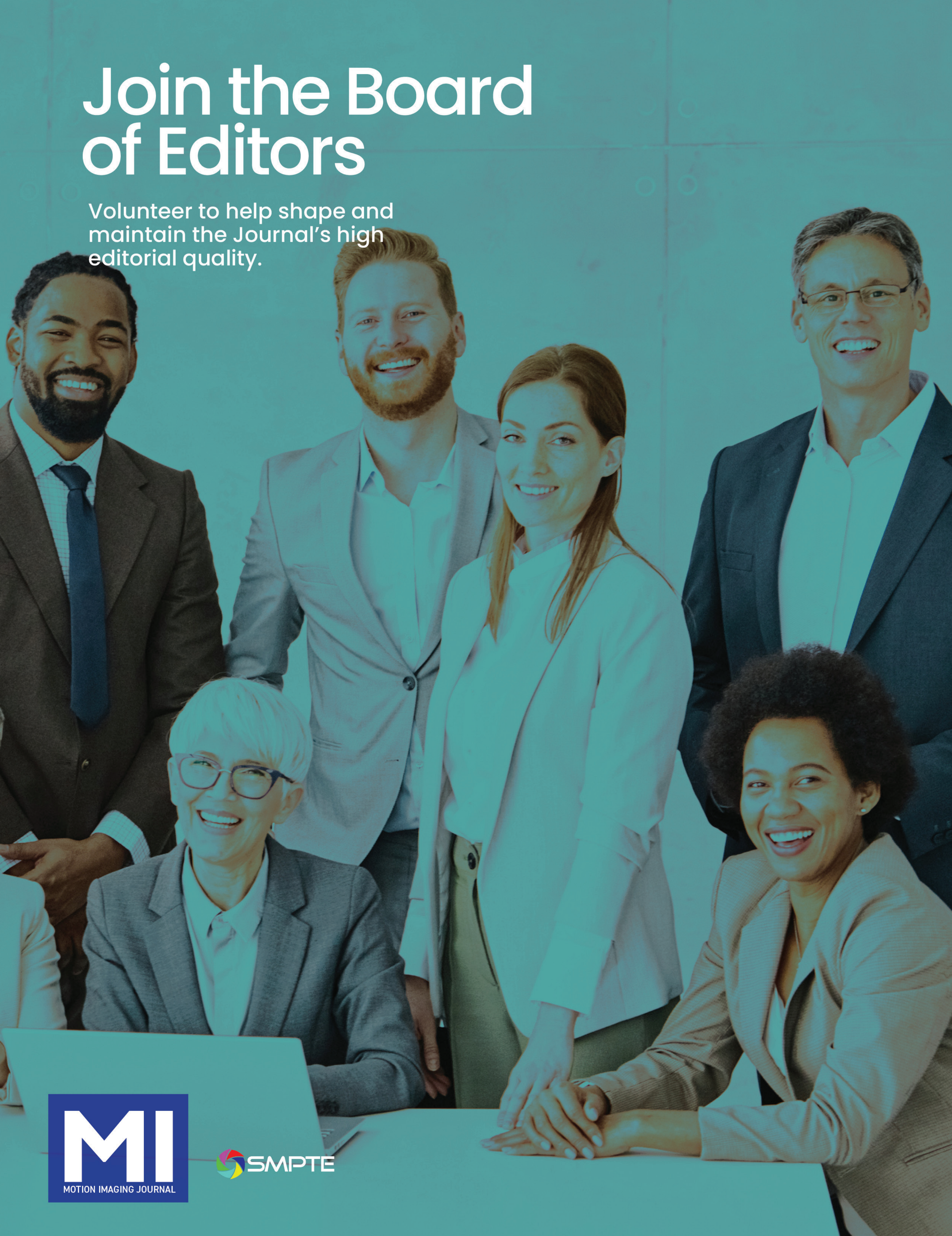
The January 1949 *Journal* published in: "Motion Picture Photography at Ten Million Frames per Second" by Brian O'Brien and Gordon Milne: "To expose a conventional motion picture at a speed of several million frames per second would require a speed of film movement of the order of 200,000 feet per second for 16-mm film, a rate entirely beyond anything attainable at present...At top rate of speed this film is driven past the image at the rate of 400 feet per second. By proper attention to mechanical and optical detail it is possible to resolve approximately 100 lines per millimeter on this film. At a resolution of 80 lines per millimeter with the film traveling 400 feet per second, the individual exposures are of one-ten-millionth-second duration, and in effect ten million separate motion picture frames per second are photographed."

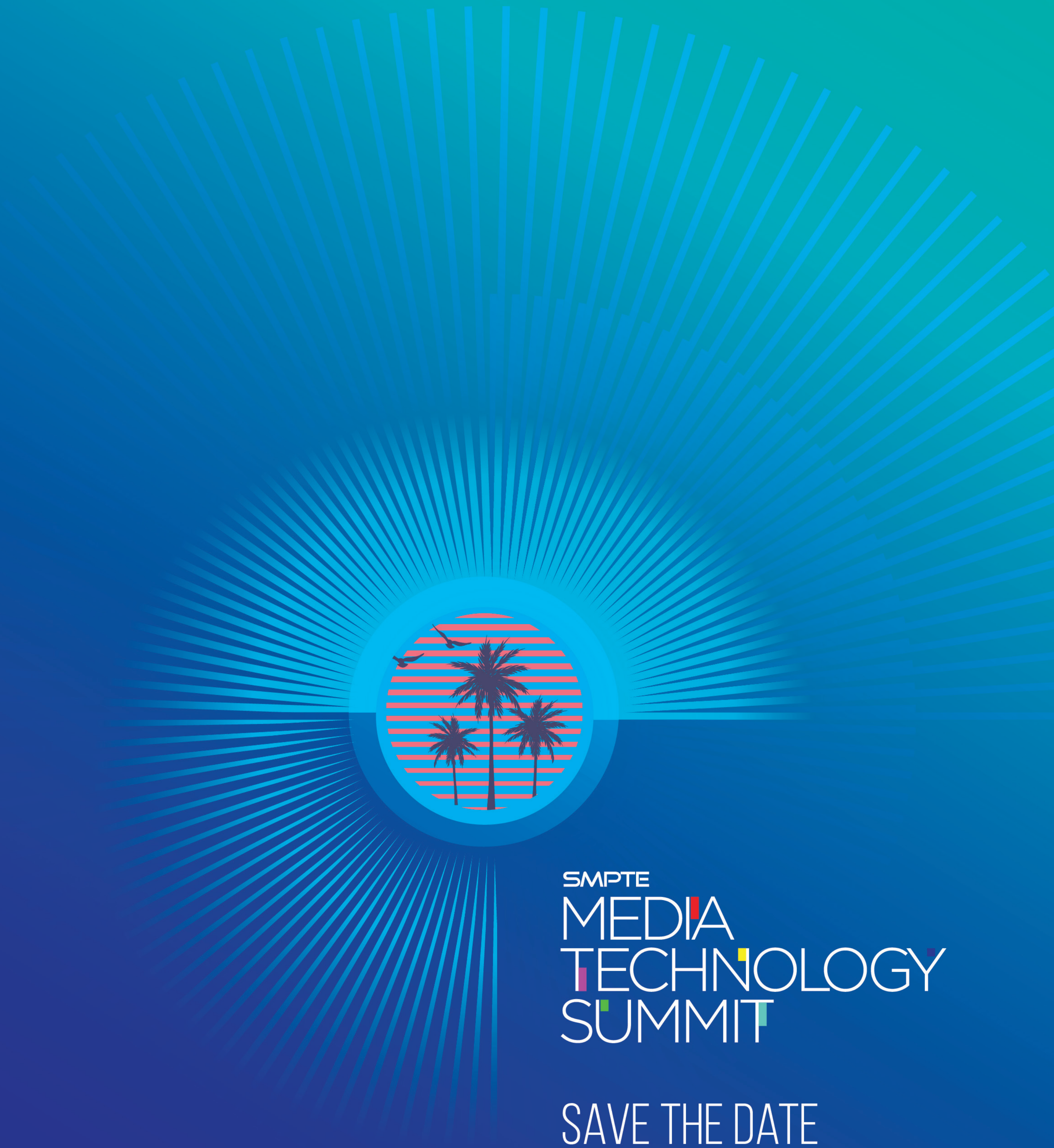
> Nineteen hundred and ninety-eight was a milestone year for the television broadcast and production community as it began the transition to DTV. The Consumer Electronics show kicked off the year with prototype receivers being displayed by almost every manufacturer. Temporary broadcast facilities were set up by ABC, CBS, and PBS to provide receiver manufacturers with actual HDTV broadcast signals.

100 Years Ago in the Journal

The May 1924 Journal published in: "The progress of Arc Projection Efficiency" by P. R. Bassett: "In the early eighteen nineties, when experimenters in all countries were struggling with the problem of making the newly found motion picture practical enough for public exhibition, it was our own founder, C. F. Jenkins, who discovered the key to the problem. The original Jenkins projector, now in the United States National Museum, consisted of an optical system which first successfully projected motion pictures on a screen. This system consisted of a slightly inclined carbon arc as the light source, two plano-convex condenser lenses to concentrate the light on an aperture plate, and an objective lens to project the image on the screen. For twenty-five years this simple and very practical arrangement held undisputed sway...The last six or seven years have seen just such an awakening of interest in methods of projection...The carbon arc must now share the field with two other light sources: the concentrated filament incandescent lamp, and the high intensity arc."

DOI: 10.5594/JMI.2024/IDYQ4827
Date of publication: 24 January 2024





SMPTE

MEDIA TECHNOLOGY SUMMIT

SAVE THE DATE

21-24 OCTOBER 2024