



# DISCLAIMER

Harris Broadcast is a global supplier of hardware and software that creates, manages, transmits, distributes and monetizes video content. Headquartered in Denver, Colorado, the company is widely recognized as the preeminent brand in the broadcast communications equipment industry. Harris Broadcast's uniquely integrated digital media solutions streamline the entire content workflow from live production to transmission. With a presence in key locations around the world, Harris Broadcast supports media customers in more than 130 countries.

On February 4, 2013, The Gores Group, a Los Angeles based investment firm that acquires controlling interests in mature and growing businesses, completed the acquisition of Harris Broadcast Communications from Harris Corporation. Harris Broadcast is working with The Gores Group to drive innovation, streamline operations and enhance customer service.

[Harris Broadcast is an independent company not affiliated with Harris Corporation.](#) This presentation and its contents may contain Harris Broadcast proprietary and confidential information. Any review, reliance, distribution, disclosure, or forwarding without expressed permission is prohibited.

Copyright ©2013 Harris Broadcast. All Rights Reserved.

# LIPSYNC AND AFD STANDARDIZED INTEROPERABLE SOLUTIONS

Paul Briscoe



# TWO TOPICS, TWO SMPTE STANDARDS

- Lipsync and Aspect Ratio Management
  - Legacy problems
  - Worse today than ever
  
- Lipsync – future standard
  
- Aspect Ratio – existing standard

## NOT THIS KIND OF LIPSYNC



Ashlee Simpson  
Lindsay Lohan  
Mariah Carey  
50 Cent  
Britney Spears  
Shakira  
Katy Perry

# NOT THIS KIND EITHER!



We're talking this kind of lipsync



## What's the big deal?

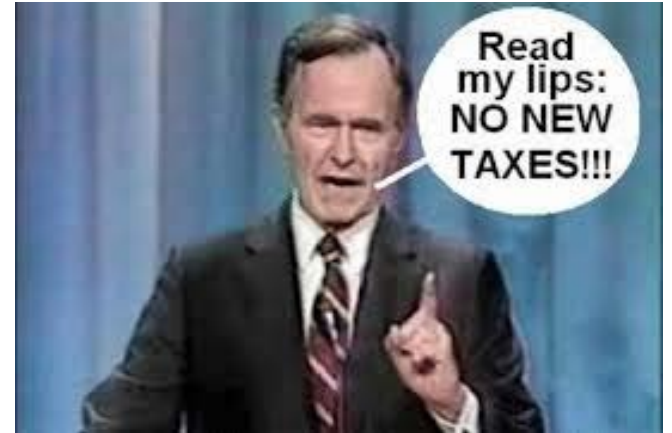
- Sound and picture aren't temporally aligned
- Audience impacts vary with content
  - Music, SFX, close-up dialogue
    - Depends on what you can see
- Direction of error has different sensitivities
  - Late is natural (1 foot =  $\sim 1$  ms)
  - Early is unnatural (no natural phenomenon)
- Human susceptibility threshold hysteresis effect
  - Once noticed, hard to miss



# WHY IS THIS A PROBLEM?

## □ Diminished Viewer Experience

- Esthetic enjoyment, QoE
- Irritation – annoying to watch and listen to
  - Change channel
  - Pay less attention or ignore
    - Advertisers don't like this!
- Loss of 'suspension of disbelief'
  - Probably why you were watching it
- Loss of believability of content
  - Advertisers really don't like this!



## It's not a new problem

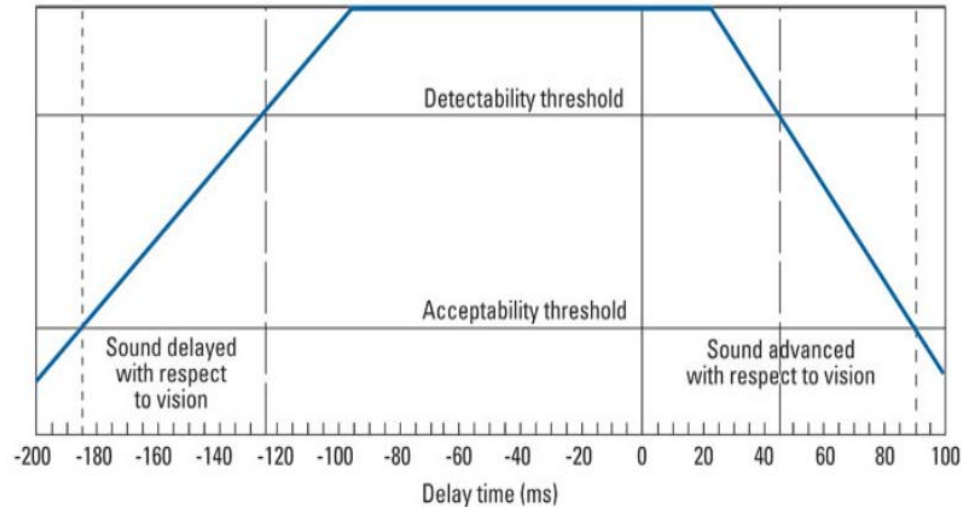
- AV Sync is as old as the motion media industry
- Silent films – piano player had to watch and keep up
- Talkies – separate sound and visual recording
- Modern film – discrete processes, post-production
- Television – not bad at first, then...
  - Postproduction
  - Digital Technology - latency
  - More Digital Technology – systems
  - Still More Digital Technology – compression
  - Internet

## EXASPERBATING FACTORS

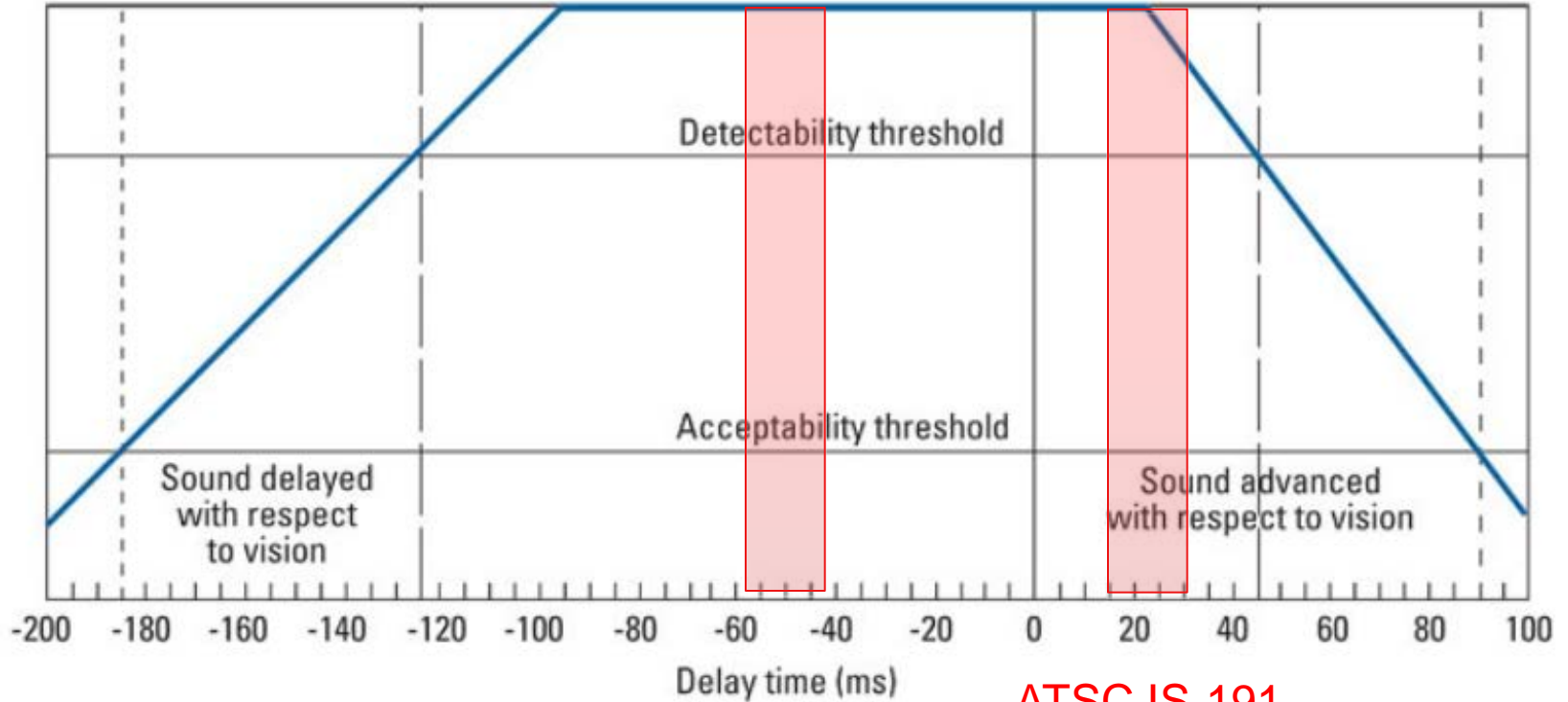
- Rich content / less formal shooting styles
- HD / larger screens
- Modern TV ‘set’ technology – complex = delay
- Complex system designs
- Use of compression – storage and transport
- Uncorrected-for signal processing
- Complex and multiple distribution architectures
- *Viewer awareness*

# THE TRIGGER PHENOMENON

- A brief trigger above detectability causes persistence  
Once you've seen it, you can't unsee it.
  - Moves one's perceived acceptability threshold
- May occur infrequently
  - Not a problem if LS good
  - Highlights bad LS
  - Can traverse content
  - Takes time to reset brain



# ITU AND ATSC THRESHOLDS

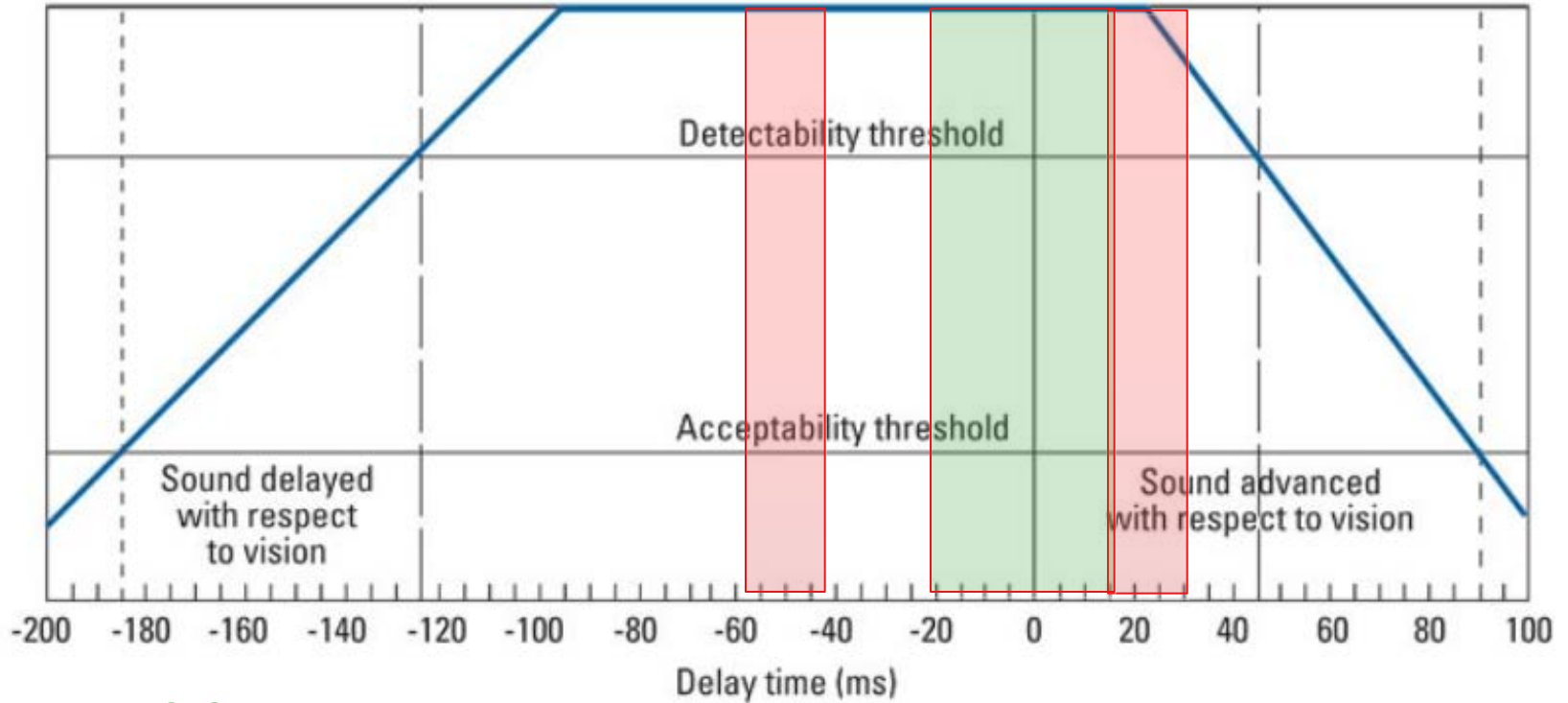


ATSC IS-191

# ORIGINAL ITU THRESHOLDS

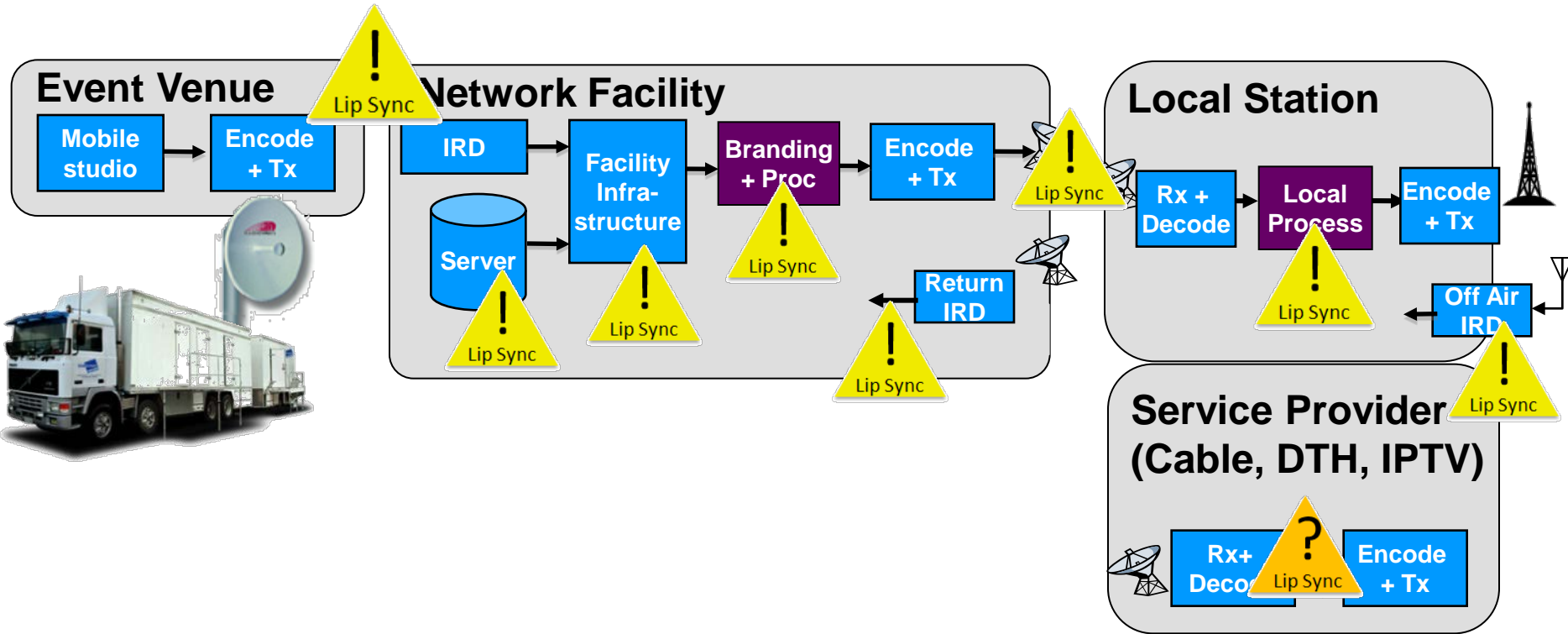
- ITU-R BT.1359
  - Early audio 45 to 125 ms
  - Late audio 90 to 185 ms
- ATSC IS-191
  - Early audio 15 ms
  - Late audio 45 ms
    - +/- 15 ms
- Film
  - +/- 22 ms

# MODERN ACQUISITION SPECIFICATION



PBS Spec

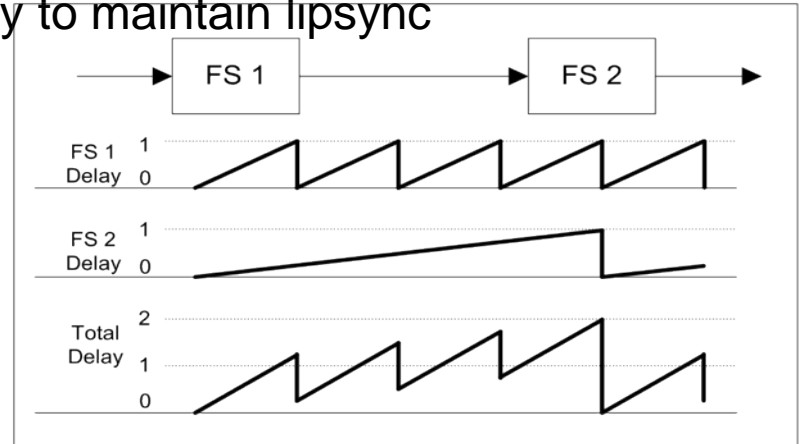
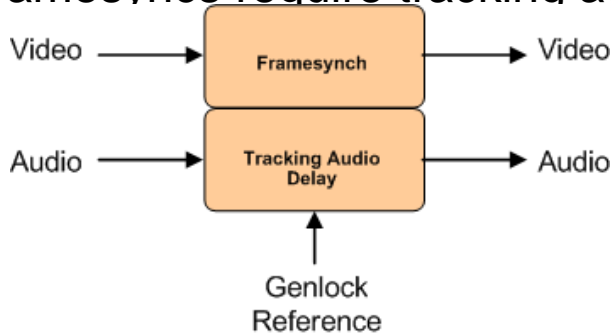
# WHERE CAN IT HAPPEN?





# EXAMPLE OF A PROBLEM SOURCE

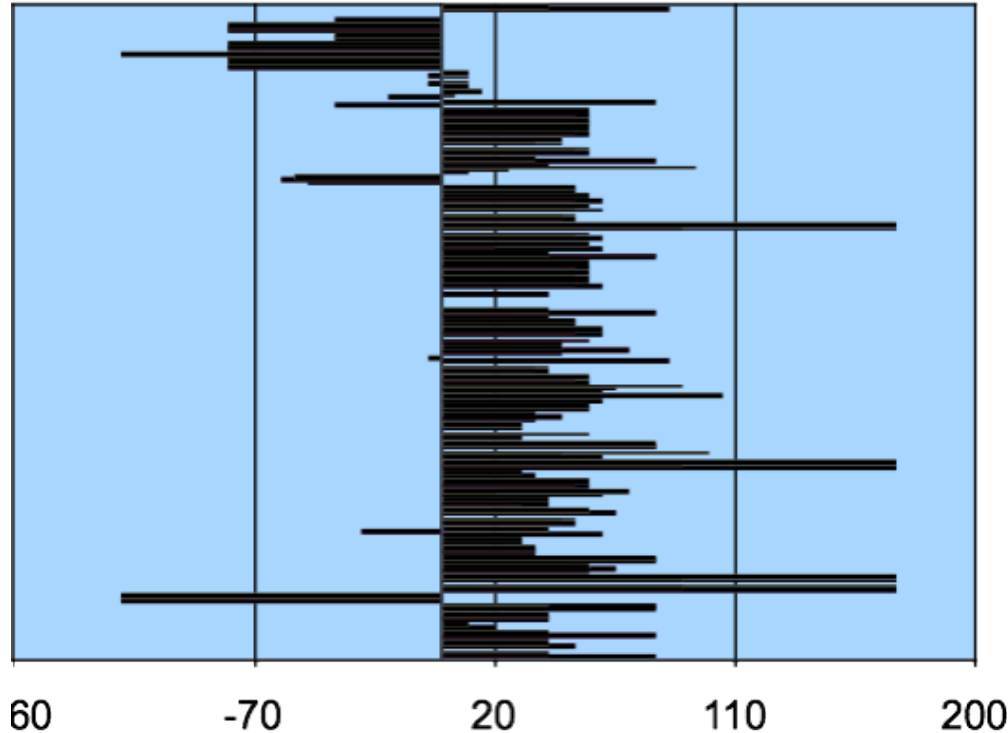
- Framesyncs ‘fix’ video timing – lock to new reference
- One framesync in the path may not be a problem
- Concatenated framesyncs can produce peak delays which can occasionally trigger the viewer
- Framesyncs require tracking audio delay to maintain lipsync



# SPREADING THE BLAME

- ❑ Vision mixers (production switchers) – DVE, reentry
- ❑ Routing switchers – FS integration
- ❑ Cameras – sensor processing latency
- ❑ Up / Down / Cross conversion
- ❑ Compression technologies – takes time on both ends
  - Contribution, storage, editing, distribution
- ❑ Viewing environment variability
- ❑ Overall, the end-to-end system from acquisition through production and distribution to final viewing provides many opportunities for AV asynchrony

# COMPARISON OF SET-TOP BOXES



## ✓ MPEG1 LII

- 41% < +/- 20 ms (past 49%)
- 52% at +/- 40ms (past 24%)
- 4% at +/- 80ms (past 24%)
- 3% > +/- 80ms (past 3%)
- max values at -160ms and +80ms

## ✓ Dolby Digital/Dolby Digital Plus

- 53% < +/- 20ms (past 44%)
- 39% at +/- 40ms (past 34%)
- 7% at +/- 80ms (past 23%)
- 1% > +/- 80ms
- max values at -160ms and +80ms

## ✓ HE-AAC

- 46% < +/- 20ms
- 34% at +/- 40ms
- 13% at +/- 80ms
- 7% > +/- 80ms
- max values at -160 and +160ms

# WHY IS IT SO BAD

- Lack of standards for the producer and broadcaster
  - Done by humans, if at all
  - No single interoperable way to measure
    - *Especially in-service*
- Lack of standards for the distributor and consumer
  - MPEG encoding standards are strict
  - *MPEG standards DO NOT specify decoding*
    - *CEA CEB-20 attempts to fix this*

# TRADITIONAL LIPSYNC MANAGEMENT

- Acquire A + V in sync
- At each stage in the chain, locally manage AV differentials manually
- Pre-offset A/V prior to ‘unfixable’ system elements
- Monitor with human QC at key points in workflow
- *Tell the viewer who calls ‘looks good leaving here, must be your set’*
  - *If they can find someone to call*
- These may not completely solve end to end problems, but can certainly minimize them
- Careful system design and equipment selection is key to *limiting* the problem

# TRADITIONAL LIPSYNC MEASUREMENT

- Most often done with human eyes
  - Subjective, variable expertise and opinion
- Only valid at time of measurement (system)
- Only valid for the content being looked at
- Automated techniques exist, but
  - Most require upstream insertion of some marker
  - Not all have system-level robustness
  - Proprietary (patents galore)
  - Not system-wide solutions
  - Some can drive correction, others don't

# Measurement today – out of service

- Simple techniques
- Can be automated
  - Many vendor solutions
- Relatively foolproof at time of use
  - If things change later, you don't know
- Human involvement
- Can't be used within content, on-air or in realtime
- Generally for acquisition, editing and system testing

# Measurement Today – in service

- Can be used on air within content
  - Various manufacturers
  - Various techniques, capabilities
  - Only for within broadcast plant
  - Varying degrees of complexity
  - May not traverse all processing
  - Box or module level solutions
  - NON-INTEROPERABLE among manufacturers
  - *Non-Standardized*
- Manufacturers
    - Dolby
    - Evertz
    - Miranda
    - Sigma
    - K-Will
    - Astro Design
    - Asaca
    - *And more*



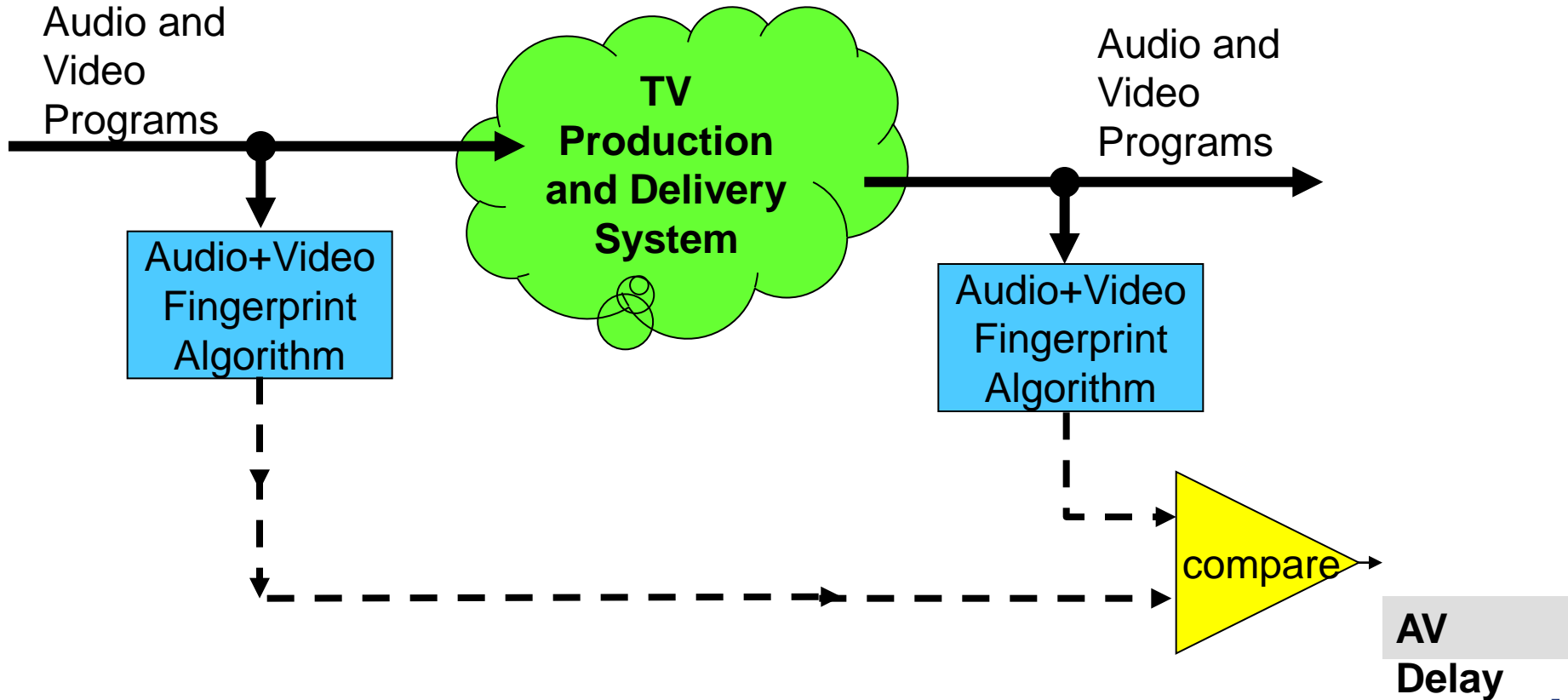
## WHAT IS NEEDED?

- Standardized in-service measurement
  - Can be used on air within live content
  - Various manufacturers
  - Same techniques used everywhere
  - Low degree of complexity – can be on a port of any device
  - INTEROPERABLE among manufacturers
  - Works throughout chain, not just in-plant
  - Must traverse all processing – *right to the home viewer*

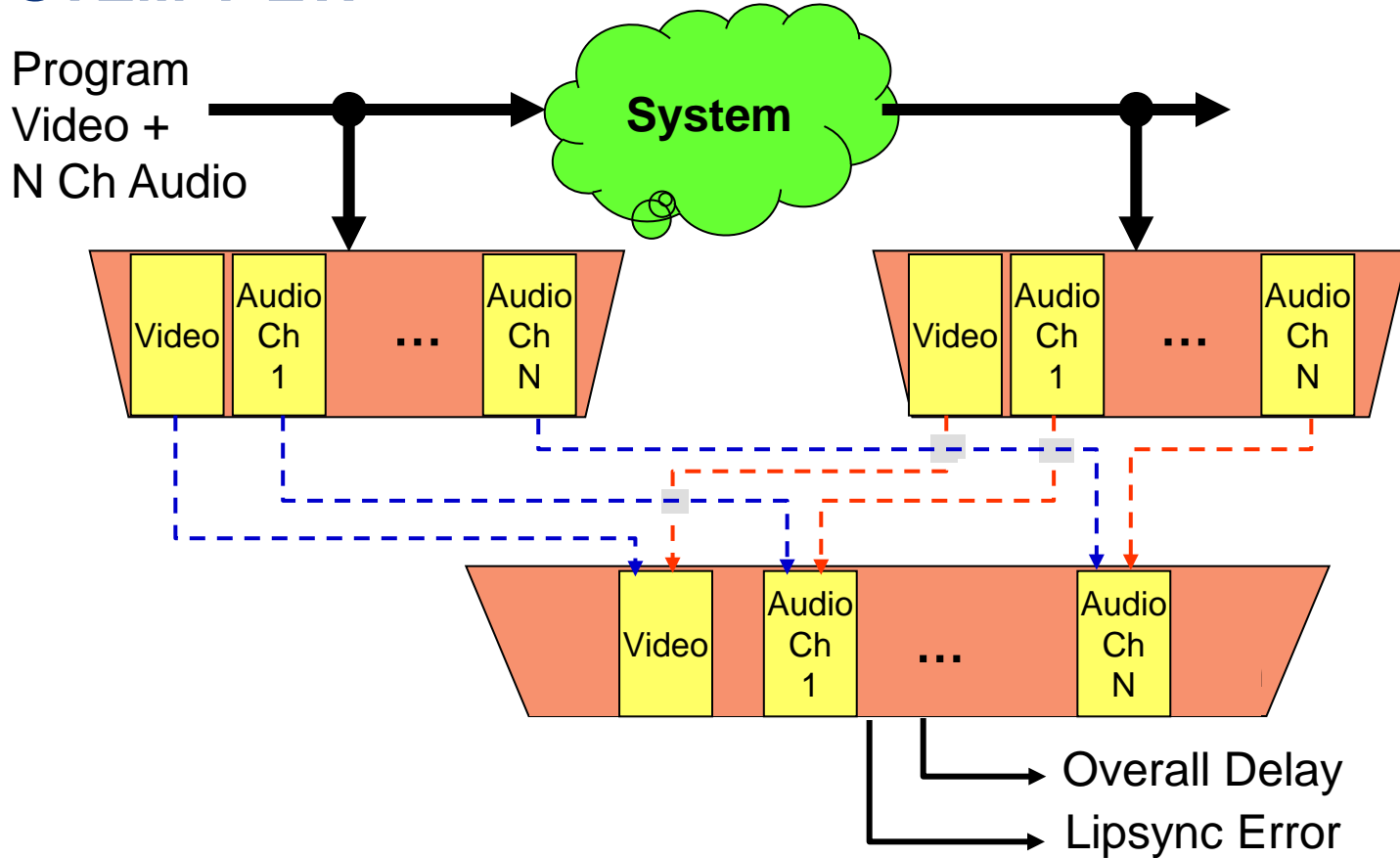
## STANDARDS TO THE RESCUE

- A standardized solution would enable interoperability
- SMPTE Standardization activity underway
- Based on **fingerprinting** the content
- Fingerprint metadata is associated with the essence
- Downstream, new fingerprints are generated
- Correlated with fingerprint metadata
- AV delay measured

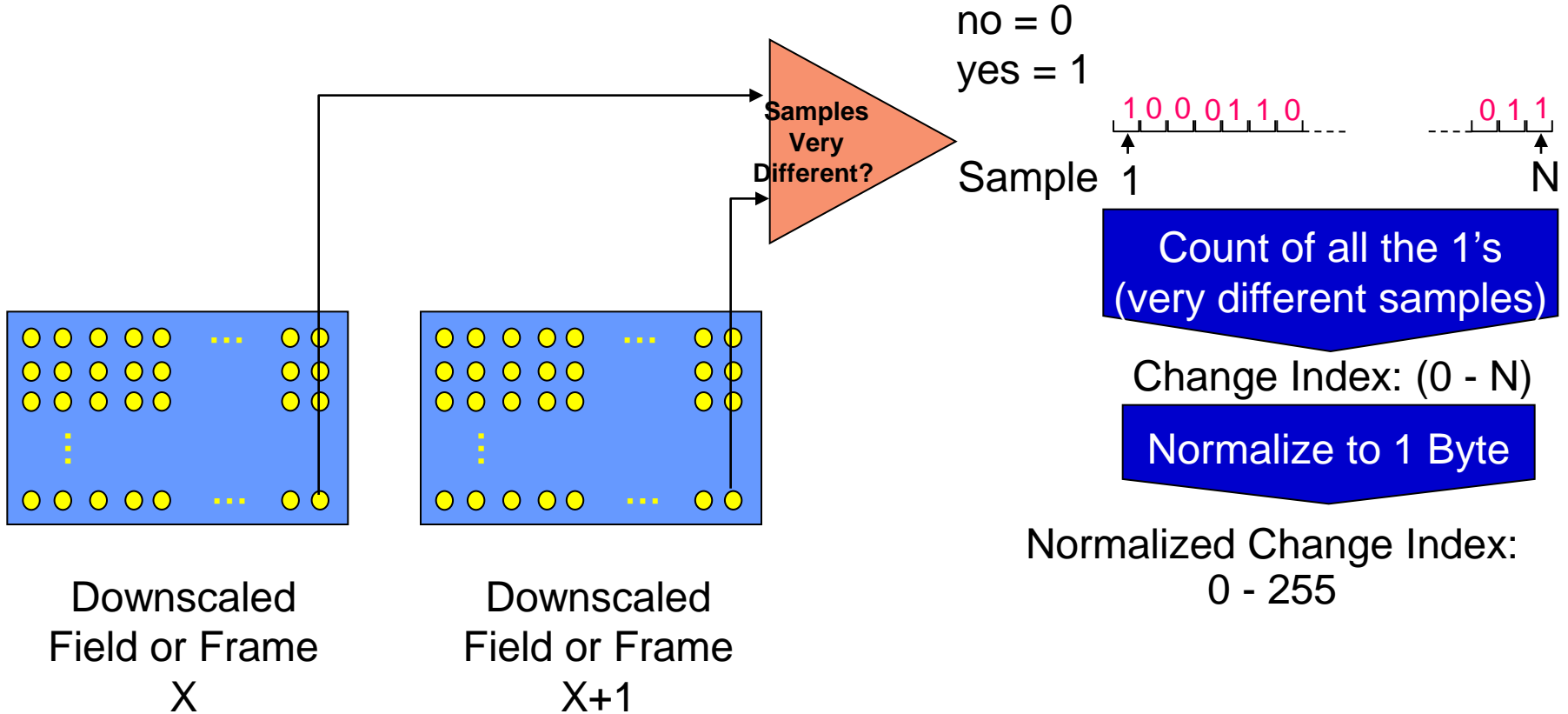
# SYSTEM VIEW



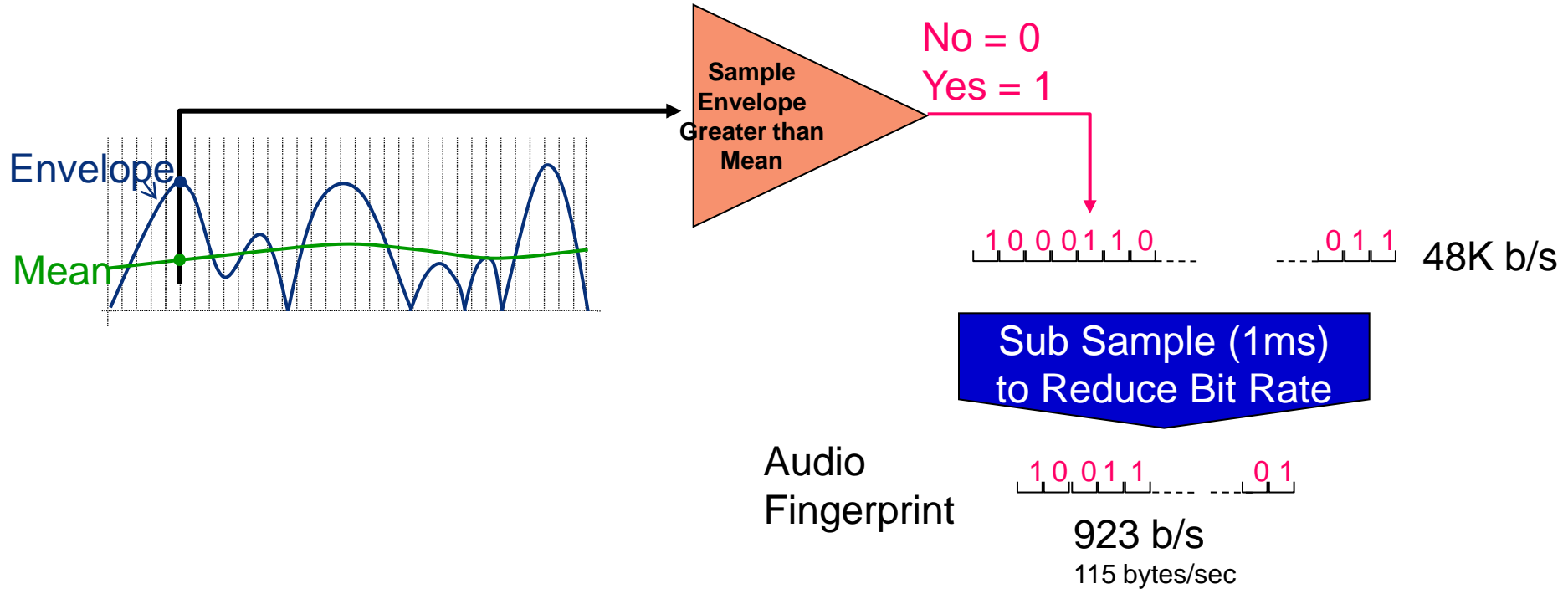
# SYSTEM VIEW



# VIDEO FINGERPRINT GENERATION



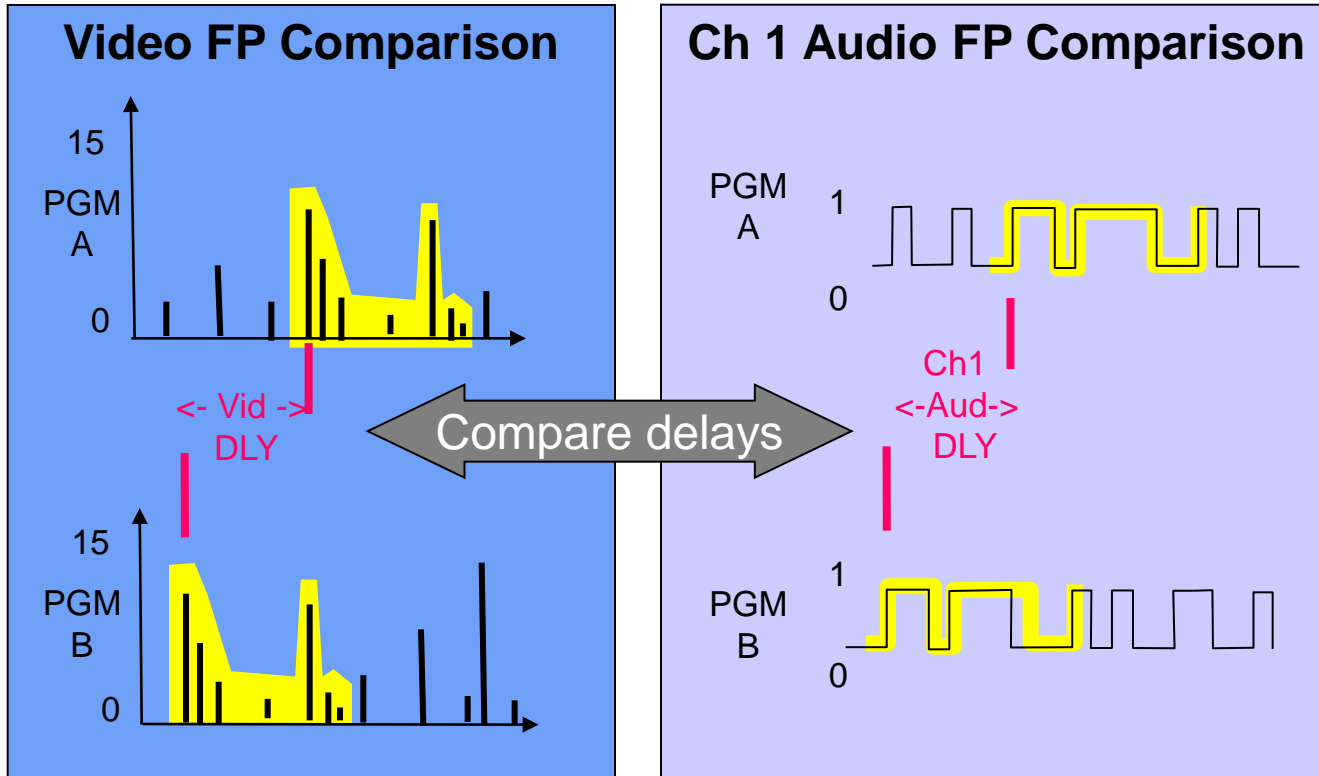
# AUDIO FINGERPRINT GENERATION



# FINGERPRINT ENCAPSULATION

- Video and Audio Fingerprints are combined into a payload container
  - Protocol Version
  - Sequence Counter
  - Video Framerate
  - Status bits – what fingerprints are present
  - ID Descriptor – future use – ISAN-type identifier
  - Video Fingerprint data
  - Audio Fingerprint data
  - Checksum

# FINGERPRINT CORRELATION





# TRANSPORT OVER SDI

- Carried in VANC packets (ST 291)
- Unique / Registered DID / SDID
- Inherently bound to the essence
  
- Essentially lossless, error-free \*see fine print



\* As much as SDI is error-free

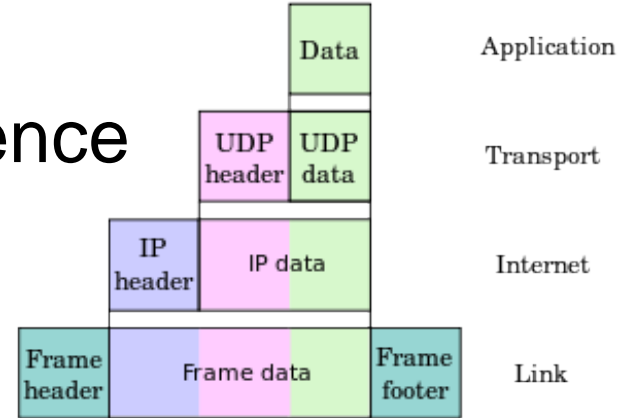
# TRANSPORT OVER MPEG TS

- Private user data in TS
- Unique PID
- Inherently bound to the essence - PAT
- Essentially lossless, error-free \* same fine print



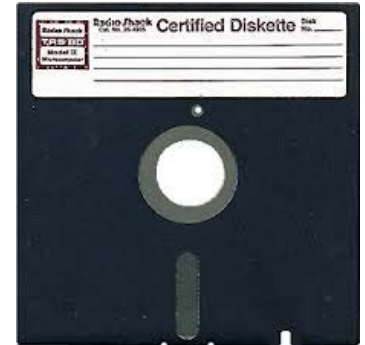
# TRANSPORT OVER IP

- UDP Packets
- More indirectly bound to essence
  - IP address
  - Payload ID
  - ID Descriptor
- May be errored / lossy / randomizing
- Requires re-association with essence at receiver



# BINDING TO FILES

- Single-part files (e.g. AVI container)
  - Embedded directly into file?
  - Associated Fingerprint file?
- Single-part files (e.g. MXF container)
  - Embedded into file package as discrete element?
  - Attached as VANC data?



# TOWARDS A STANDARDIZED SOLUTION

- SMPTE Working group on Lip Sync
- Mandate – establish technology and draft a Standard:
  - Fingerprinting algorithms for audio and video
    - Must meet performance requirements
  - Transport specifications
    - SDI, MPEG TS, IP
  - Binding specifications
    - Files

# THE DRAFT STANDARD

Group is called **24TB-01 AHG LipSync**

Standard will be called ST-2064, in multiple parts

- Part 0 - Overview of Standard
- Part 1 - Fingerprint generation
  - Payload composition
- Part 2 - Fingerprint transport – SDI, MPEG, IP
- Part 3 - File binding

# SMPTE ACTIVITIES

- Complete Part 1 Draft Editing – Q2 / 13
  - Review, address comments, elevate to pre-ballot
- Complete Part 2 Drafting – Q3 / 13
  - Draft, edit, review within group
  - Review, address comments, elevate to pre-ballot
- Formal Ballot on Parts 1 and 2 – Q4 / 13
- Begin Part 3 Drafting – Q4 / 13
- Liaison follow-up with CEA, EBU, ITU



# BEYOND SMPTE

- Method can work all the way to the home
  - IRDs
  - STBs
  - TV Sets
  - PCs
  - Web devices
- Additional transport may need to be defined
- Outside SMPTE scope - liaison



## SUMMARY

- Fingerprinting technique selected for SMPTE standardization
- Can span from acquisition / post right to the viewer's TV
- Live in-service use
- Simple to implement
- Robust
- **Interoperable!**



*Coming soon to lipsync errors near you.*

# AFD – what's the problem here?



1.33:1 (4:3)



1.56:1 (14:9)



1.67:1 16mm



1.77:1 (16:9)



1.85:1



2.35:1



1.33:1 (4:3)



1.56:1 (14:9)



1.67:1 16mm



1.77:1 (16:9)



1.85:1



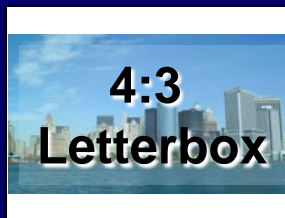
2.35:1

# WITHOUT ASPECT RATIO MANAGEMENT

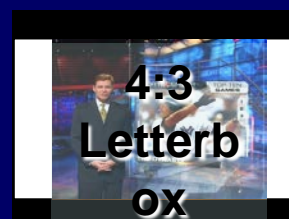
## Produce



## Broadcast



## Viewer Display



# WHAT WE REALLY WANT

## Produce



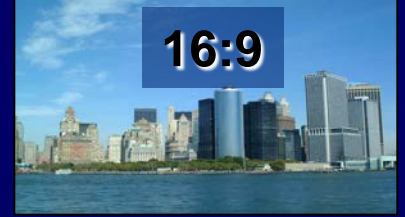
## Broadcast



## Viewer Display



## What We Want



## AFD – what's the problem here?

- Broadcasters are faced with a double-ended problem
  - Source content may be in 4:3 or 16:9
    - Archive content is 4:3
    - Network may feed one or the other
    - Station evolutionary transition to HD
      - Live Production studio still SD
      - News studio upgraded to HD
      - ENG in both formats
      - Commercials and programs are a mix
  - May have to simulcast both formats  
How do you satisfy both types of viewer?  
AFD, of course.

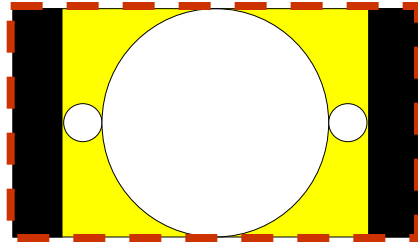
# WHAT DO WE NEED

- A method of conveying the native aspect ratio of the picture essence with respect to its “carrier” raster
  - 16 x 9 inside 16 x 9
  - 4 x 3 inside 4 x 3
  - 4 x 3 inside 16 x 9
    - Various flavors
  - 16 x 9 inside 4 x 3
    - Various flavors

# WHAT IS AFD?

- Literal Translation AFD = **A**ctive **F**ormat **D**escription
- Metadata that describes the **Format** of a signal
- May be carried with signal during production, distribution and emission
- AFD specifies the carrier frame aspect ratio and characteristics of the active image area
- AFD Metadata is carried in the Vertical Ancillary data space of the signal
- AFD is designed to automatically control aspect ratio conversion in downstream devices

# The Carrier (“coded Frame”)

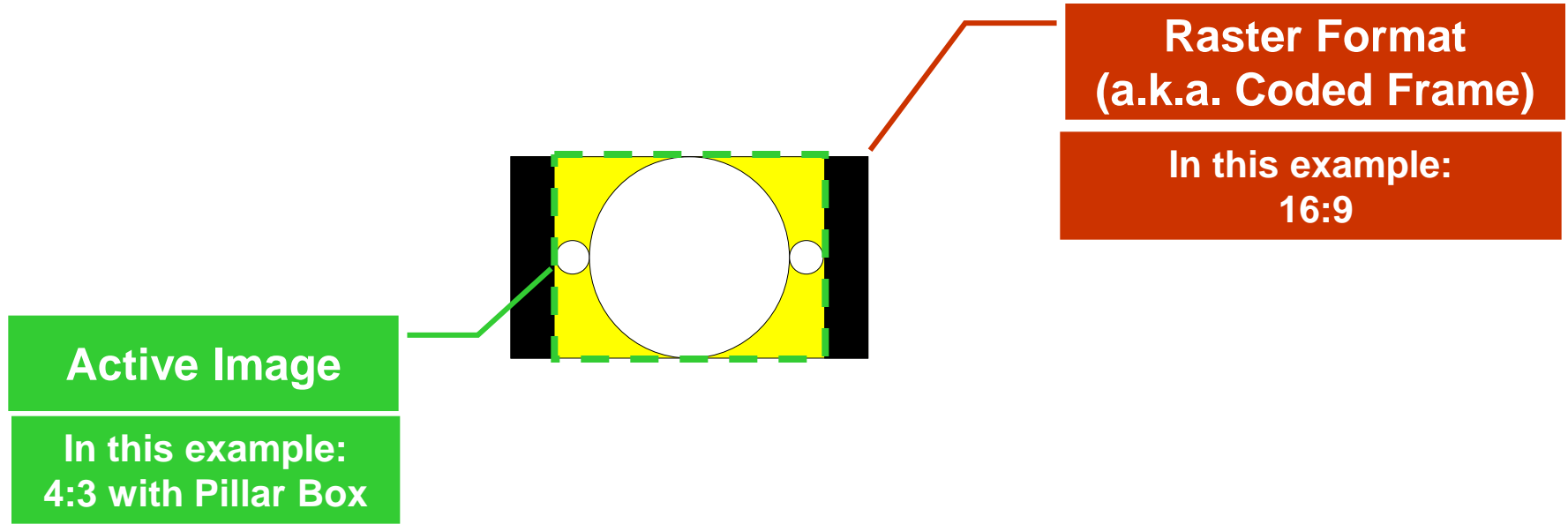


**Raster Format  
(a.k.a. Coded Frame)**

**In this example  
16:9**

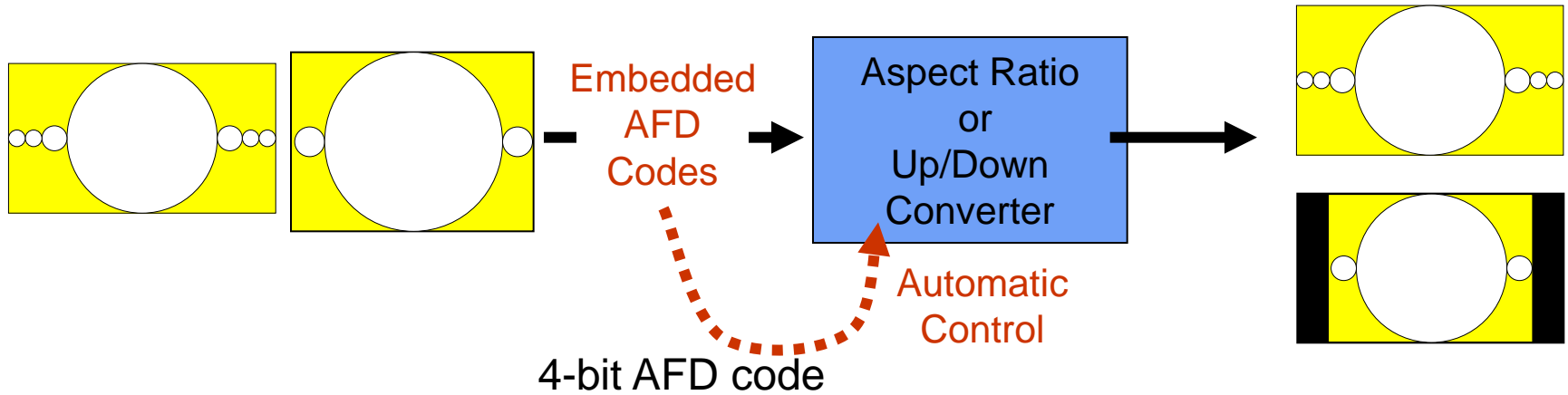


# The essence (“Active Image”)



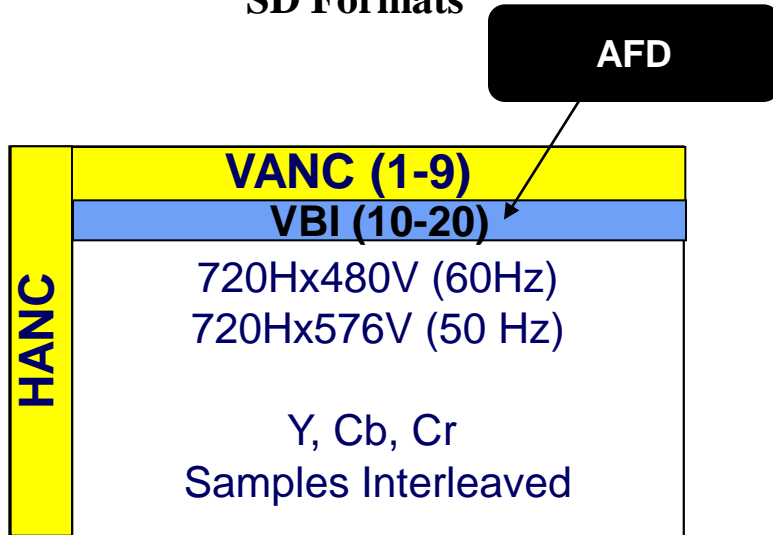
The SMPTE-2016-1 AFD specifies both the **Raster Format** and the **Active Image format** (image inside the raster )

# CONVEYING THE ASPECT RATIO

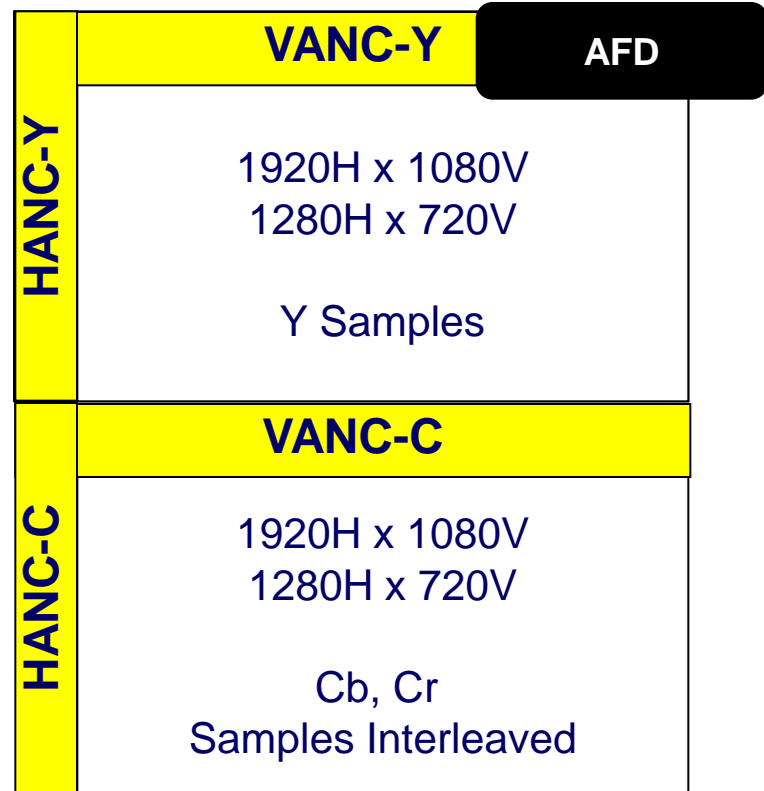


# WHERE THE AFD METADATA GOES

## SD Formats



## HD Formats



## What AFD is – and isn't

- AFD tells you
  - The format of the image inside the frame
- AFD does *not necessarily* tell you
  - What to do with the image
  - SMPTE-2016 describes the common mappings
  - Implies how to handle the picture when scaling
  - May or may not be what you want
    - Has to be correct to begin with
    - Must be updated anytime scaling is performed
  - You may elect to use codes in a non-standard way
    - Careful! Private usage won't interoperate outside your system.

## What AFD is – and isn't

- *It DOES NOT mean AUTOMATIC!*
  
- **Active** means that it's a dynamic thing that changes with the content
  
- It doesn't work without you:
  - Understanding of the format of the content
  - Understanding of the artistic intent
  - Understanding of your devices
  - Understanding of your system behavior
  - Understanding of what you want to deliver
  - Understanding of what may happen downstream.

# AFD CONTROLS ASPECT RATIO CONVERSION (ARC)

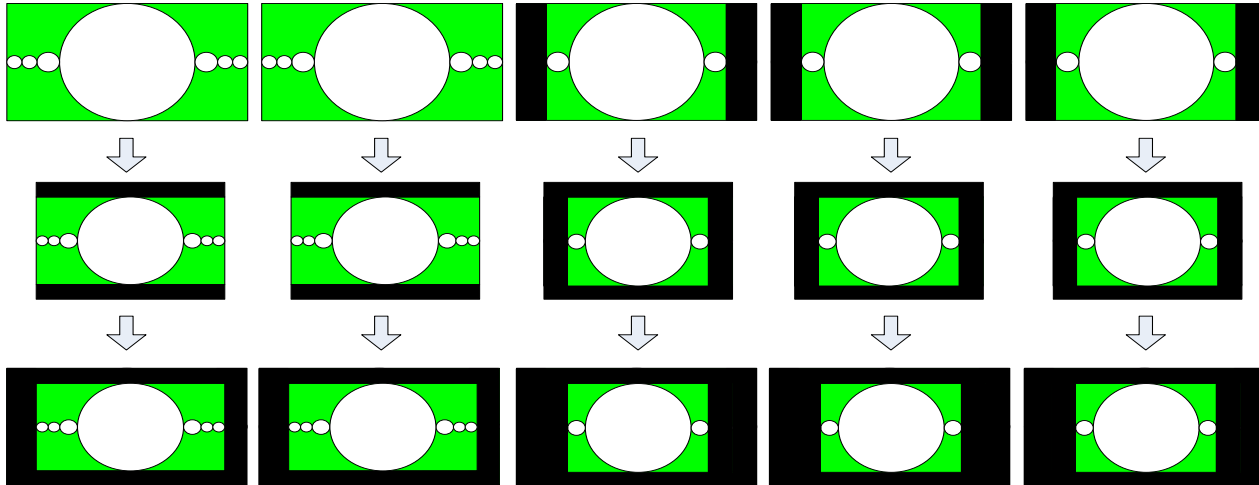
- AFD controls downstream ARC behaviour
  - Getting it wrong in one place has a downstream ripple effect
- Down and Upconversion are lossy processes
  - Concatenation of ARCs causes progressive degradation
  - $16:9 > 4:3 > 16:9$  means
    - Decimation of pixels to lower resolution
    - Interpolation of pixels back to higher resolution
  - Lost information cannot be restored
  - Image detail enhancement will not bring it back, only generate artifacts downstream
- Important to minimize amount of ARC

## What AFD is – and isn't

- AFD tells you
  - The format of the image inside the frame
- AFD does *not necessarily* tell you
  - What to do with the image
  - SMPTE-2016 describes the common mappings
  - Implies how to handle the picture when scaling
  - May or may not be what you want
    - Has to be correct to begin with
    - Must be updated anytime scaling is performed
  - You may elect to use codes in a non-standard way
    - Careful! Private usage won't interoperate outside your system.

# Downconversion – No AFD

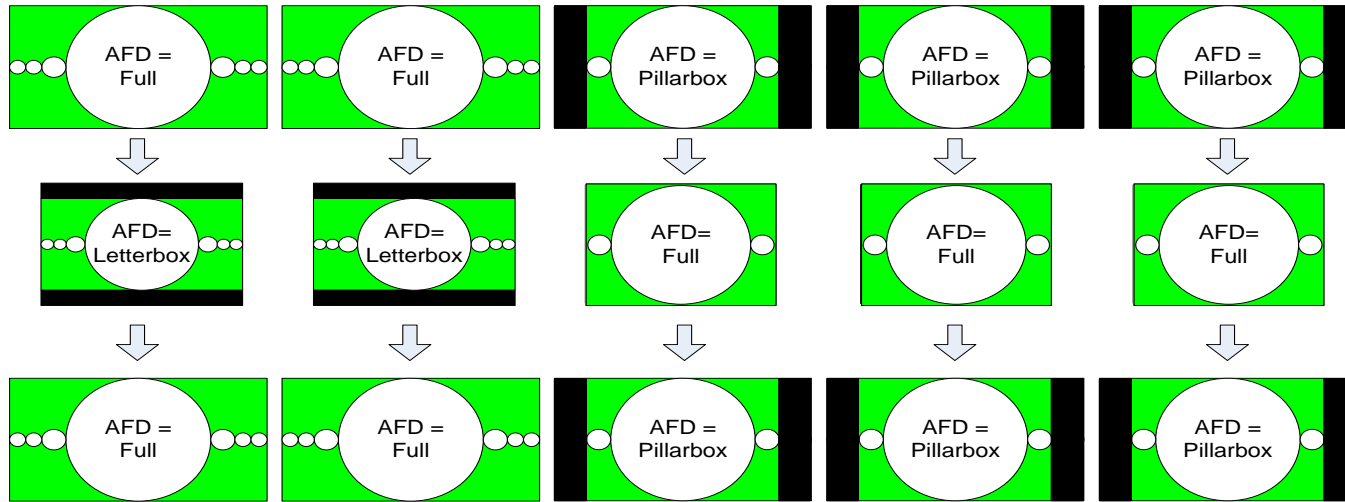
- Stream switch from HD full to SD pillarbox
- Convert from HD to SD and back





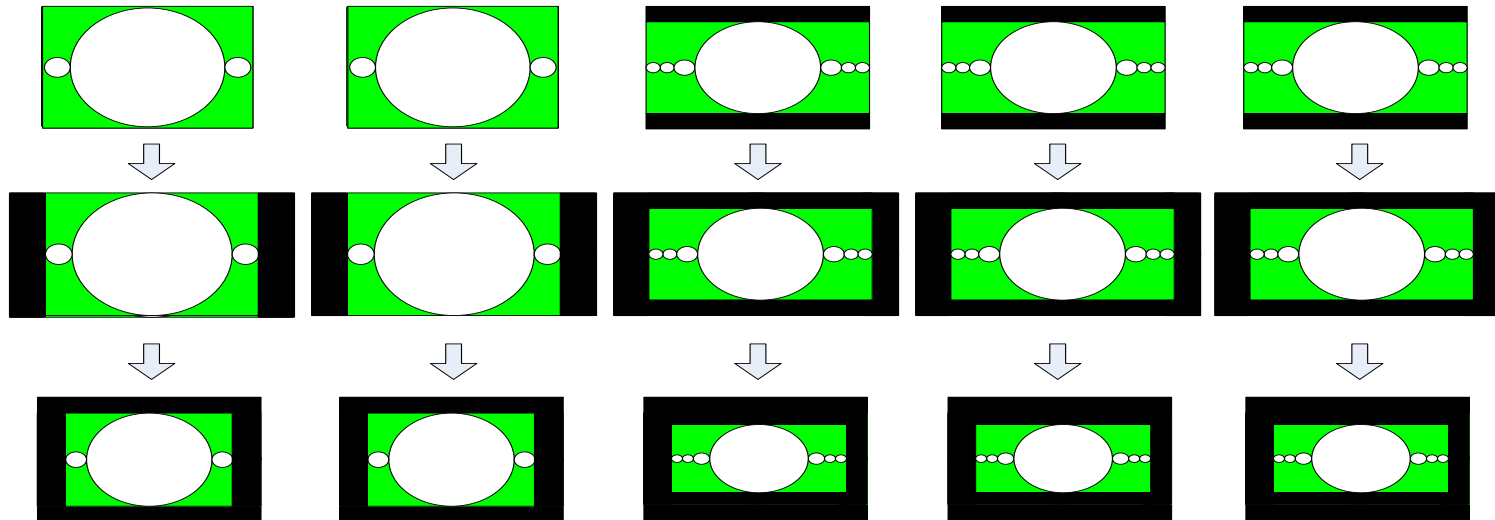
# DOWNCONVERSION WITH AFD

- Stream switch from HD full to SD pillarbox
- Convert from HD to SD and back



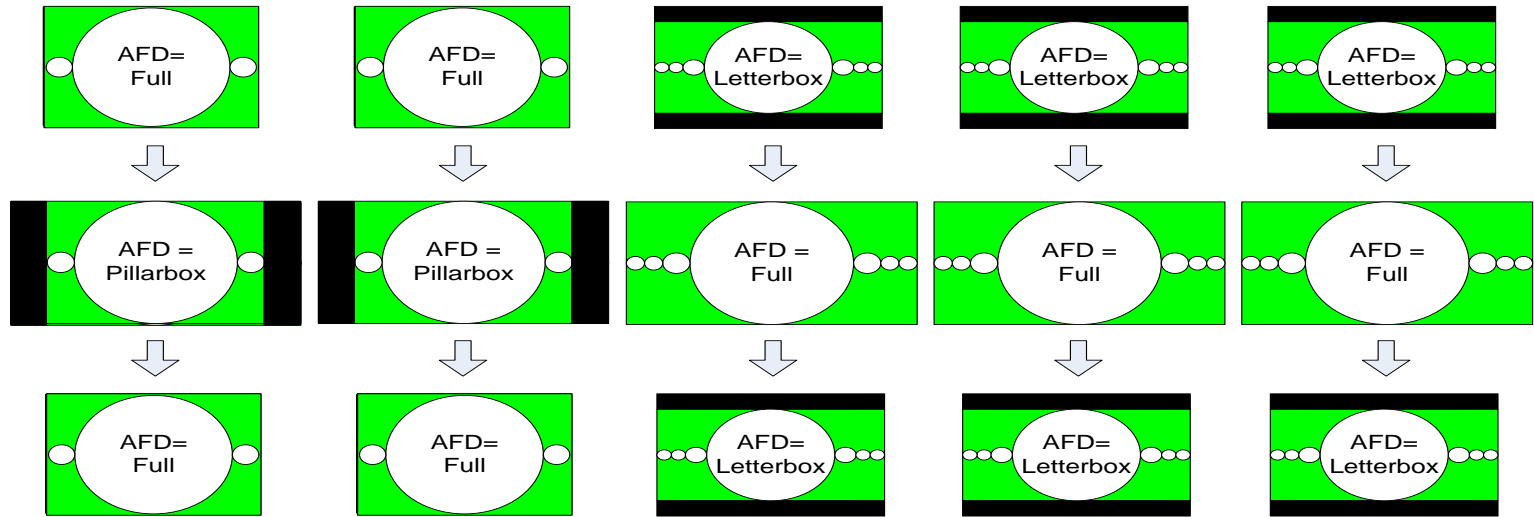
# Upconversion – No AFD

- Stream Switch from SD to HD letterbox
- Convert from SD to HD pillarbox and back



# Upconversion – with AFD

- Stream Switch from SD to HD letterbox
- Convert from SD to HD pillarbox and back



## AFD BINDING TO THE ESSENCE

- AFD is attached to every frame / field
  - Carried in VANC data in SDI
- Devices must maintain tight association
  - Non-scaling devices – leave it alone
  - Scaling devices – update it correctly
- Must be updated frame accurately
- Placement is important
  - Correct VANC space – per SMPTE Standard
  - Both fields in interlaced systems
- In progressive systems, ARC / AFD changes must occur on correct –p frame for downconversion to -i

# DOES AFD SOLVE THE ARC PROBLEM?

- It can in a broadcast system that:
  - Contains carefully qualified equipment suitable for the task
    - Live streams with dynamic AFD changes require realtime AFD / ARC handling
    - Fixed-format streams can use non-realtime or static ARC devices
  - Has content paths that are qualified for required response to AFD
    - End to end accuracy and timeliness
    - Behavior in unexpected circumstances
      - Conscious choice of failure mode behavior

# WHAT ABOUT THE VIEWER?

- This is out of our control
  - They have too many knobs to turn
  - May select a forced / manual display ARC
    - “Gotta light all those pixels I paid for!”
  - Delivery of AFD / ARC control inconsistent and incomplete at this point in time
    - Works on some channels
    - Not on others
    - May even break on some
  - Until we get consistent delivery of ARC data to the set, the problem will remain
    - Good news is it's getting closer all the time.

## WHERE DOES IT END?

There will always be opportunity for ARC escapes

- Legacy content needs to be ingested and tagged correctly

- Mis-processed content needs to be fixed

*Despite our best efforts, viewers may still override it*

It will probably never be perfect, but with care in device selection and system design, we can mitigate most opportunities for problems.

# SMPTE STANDARDS RELATED TO AFD

## SMPTE-2016

**SMPTE-2016-1** Format for Active Format Description and Bar Data  
*Specifies the AFD codes*

*This is the content spec, it explains the bits*

**SMPTE-2016-2** Format for Pan-Scan Information

*Specifies the Pan-Scan codes*

**SMPTE-2016-3** Vertical Ancillary Data Mapping of Active Format Description and Bar Data

*Defines where AFD Meta data is placed in ANC*

*This is the transport spec, it specifies where to put the bits*

**SMPTE-2016-4** Vertical Ancillary Data Mapping of Pan-Scan information

*Defines where Pan Scan Meta data is placed in ANC*

**SMPTE-2016-5** KLV Coding for Active Format Description, Bar Data, and Pan-Scan Information

*Specifies the Key values for AFD, Bar Data and pan and scan within the SMPTE MXF dictionary*



# INTEROPERABILITY

Standardization enables interoperability

SMPTE Standards are at the core of our infrastructures

AFD standard is widely deployed and works well.

Continued adoption and usage

Lipsync standard offers the opportunity to do the same

# QUALITY OF EXPERIENCE

Lipsync and AR errors hugely impact QoE

Enjoyment

Attention

Viewers change channel (or worse)

*Advertisers don't like this!*



*AFD and Lipsync standards can vastly improve QoE*

# QUESTIONS?

Thank-you!

For further information, comments, questions, flames?  
I'm happy to help, just drop me a note.

[paul.briscoe@harrisbroadcast.com](mailto:paul.briscoe@harrisbroadcast.com)



**Harris  
Broadcast**

Deliver the moment.

