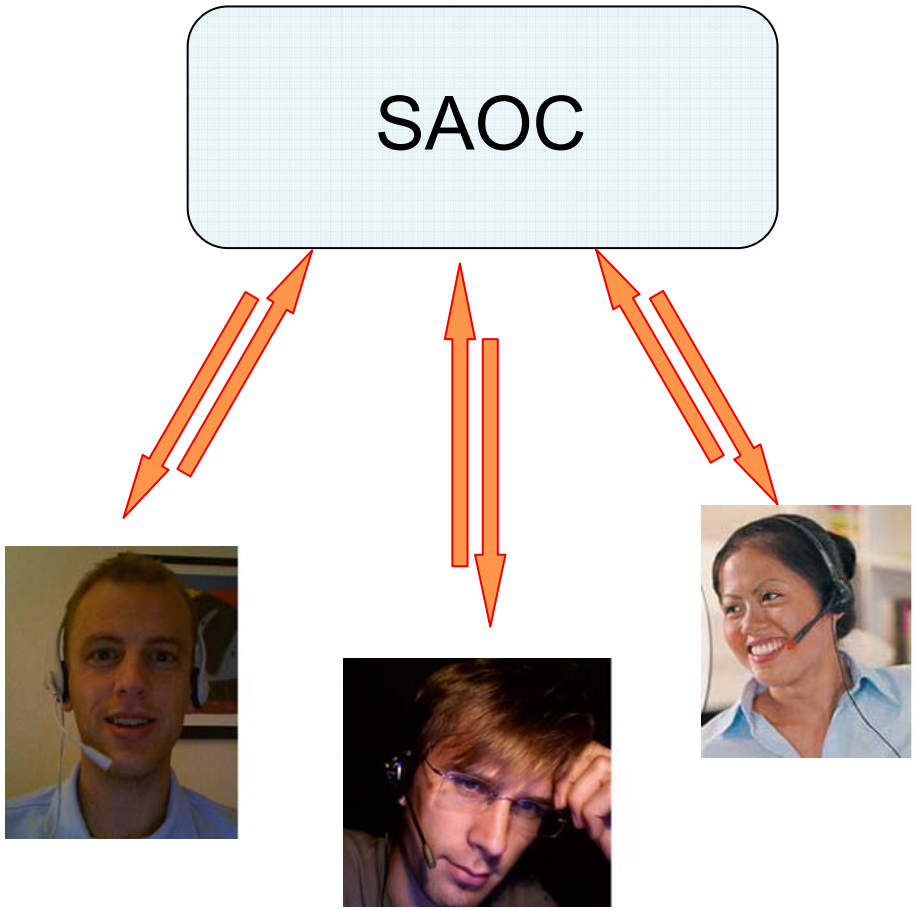
Spatial Audio Object Coding (SAOC)

- Application scenarios
- General concept
  - Transcoding object cues into MPEG Surround cues
- Technical enhancements - Part 1
  - Stereo downmix mode

Barbara Resch

Coding Technologies

Applications - Teleconferencing



- Binaural playback configuration
- Spatial positioning of the speakers

⇨ easier to recognize who is speaking

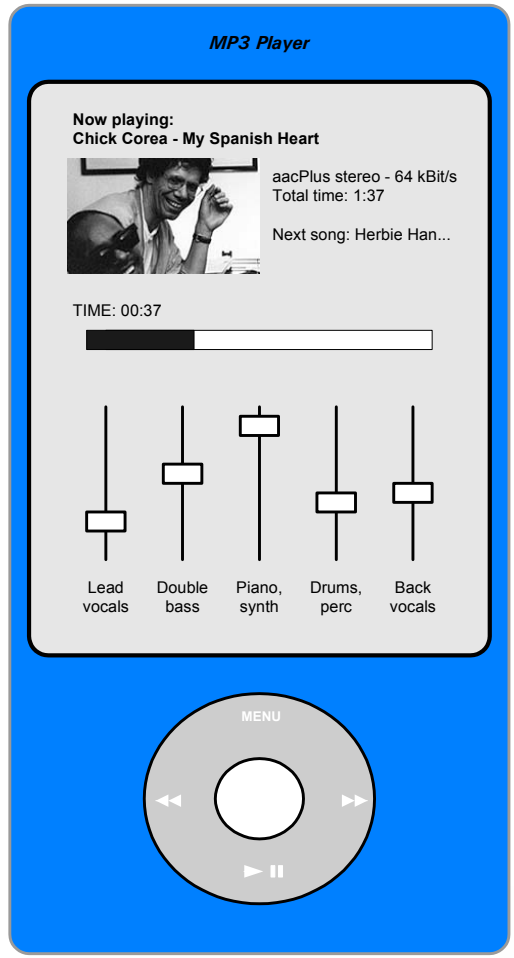
Applications - Interactive re-mix

Re-mix Music:

- “Do your own mix“
  on portable player or PC
- Karaoke (remove vocals)
- Play-along (remove instrument)

Re-mix Movies, TV:

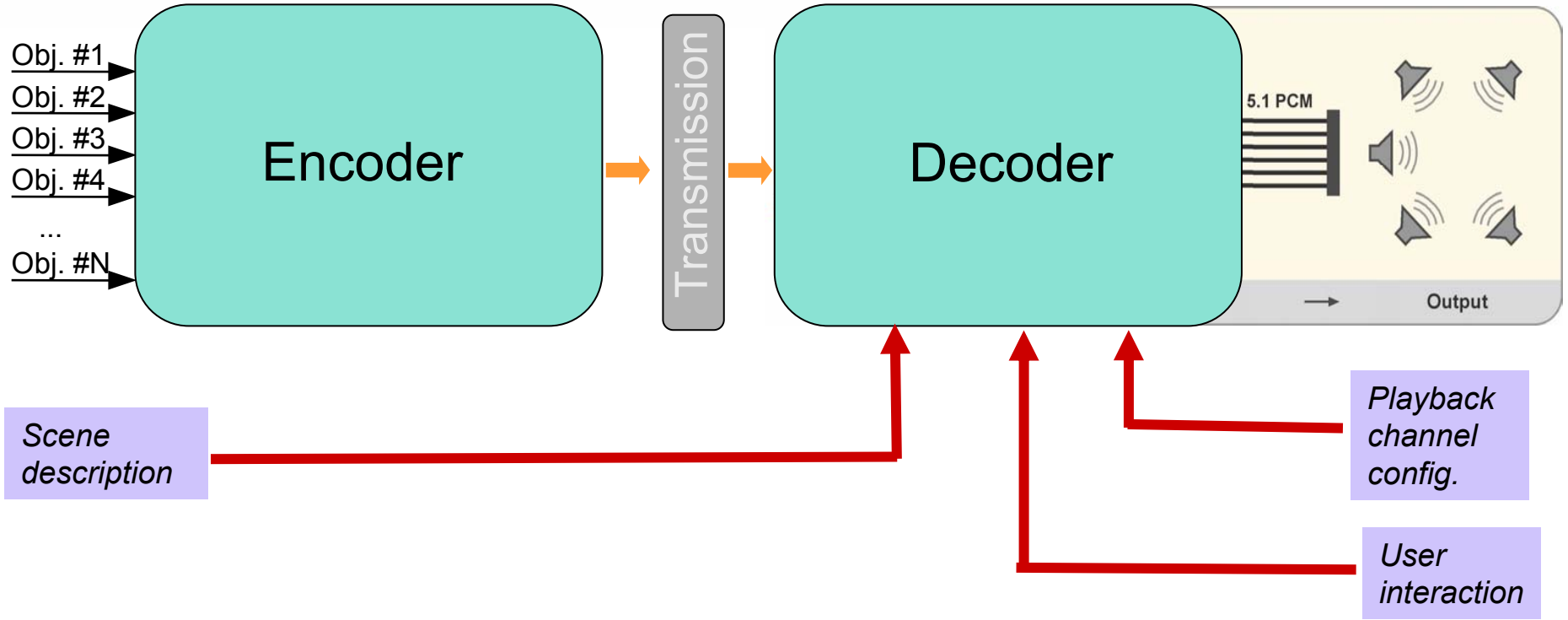
- Boost dialogue over
  background music
  (speech enhancement)



Applications - Gaming / rich media

- Game audio rendering
  - Remote players
  - Adaptive audio scene
- Virtual Reality scenes
- Animations (“Flash”-movies) also with user interaction

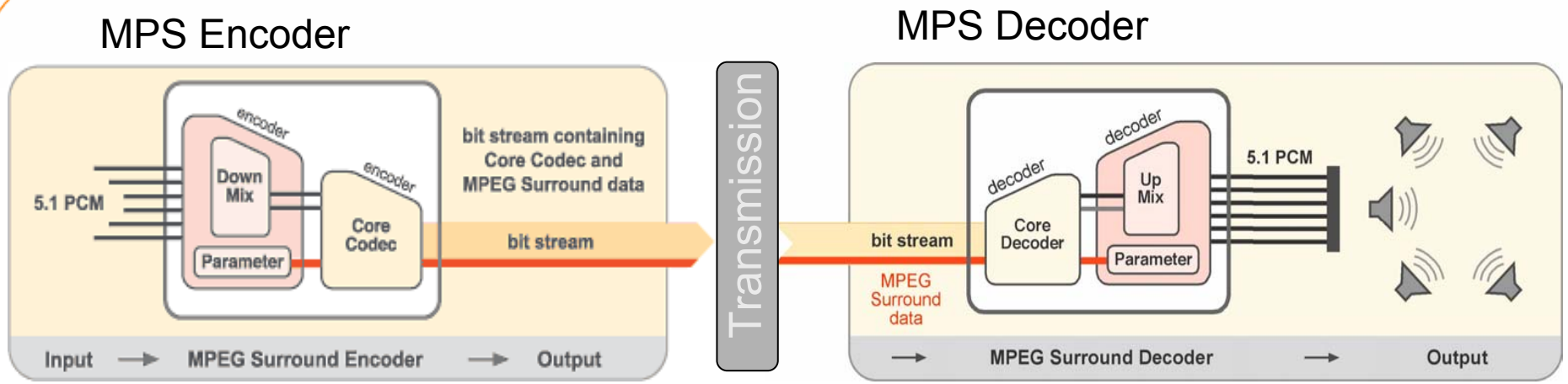
General concept - SAOC



Spatial Audio Object Coding (SAOC)

- Flexible interactive rendering of objects

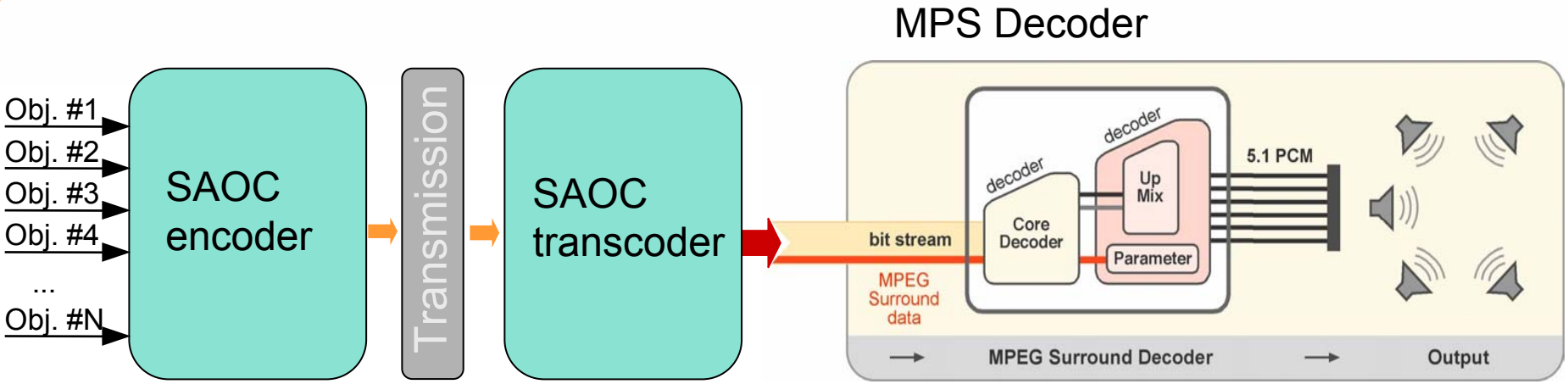
General concept – MPEG Surround (MPS)



SAOC as an extension to MPEG Surround (MPS)

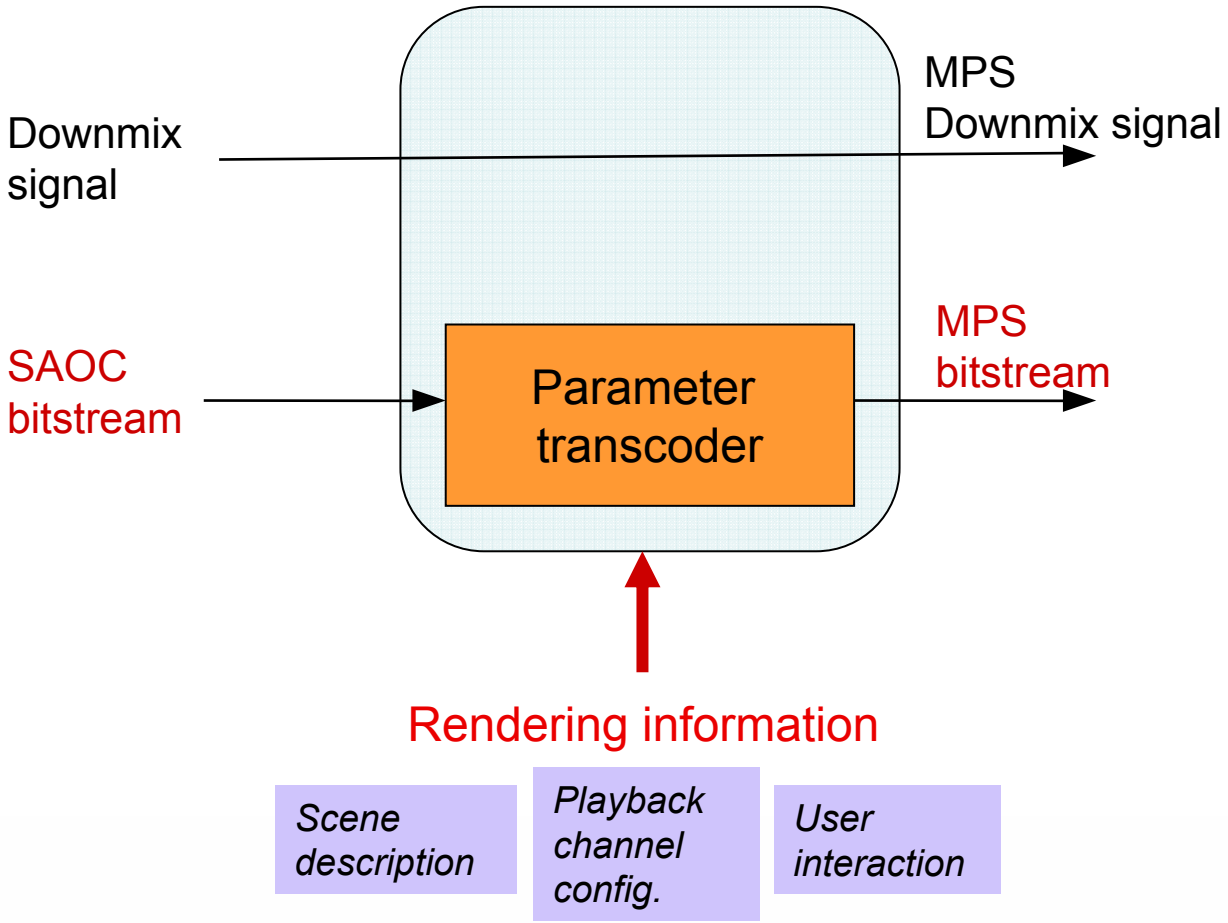
- State of the art in multi-channel audio compression
  (Standardized in 2006, ISO/IEC 23003-1)
- Covers all playback cases (binaural, 5.1, 7.1, up to 32 channels)
- Parametric description instead of coding all channels

General concept – Transcoding to MPS



- Flexible interactive rendering of objects
- Maximum advantage of existing standardized MPEG Surround technology
- Transcoding object cues into MPS cues
- Different playback cases

SAOC transcoder: Mono downmix

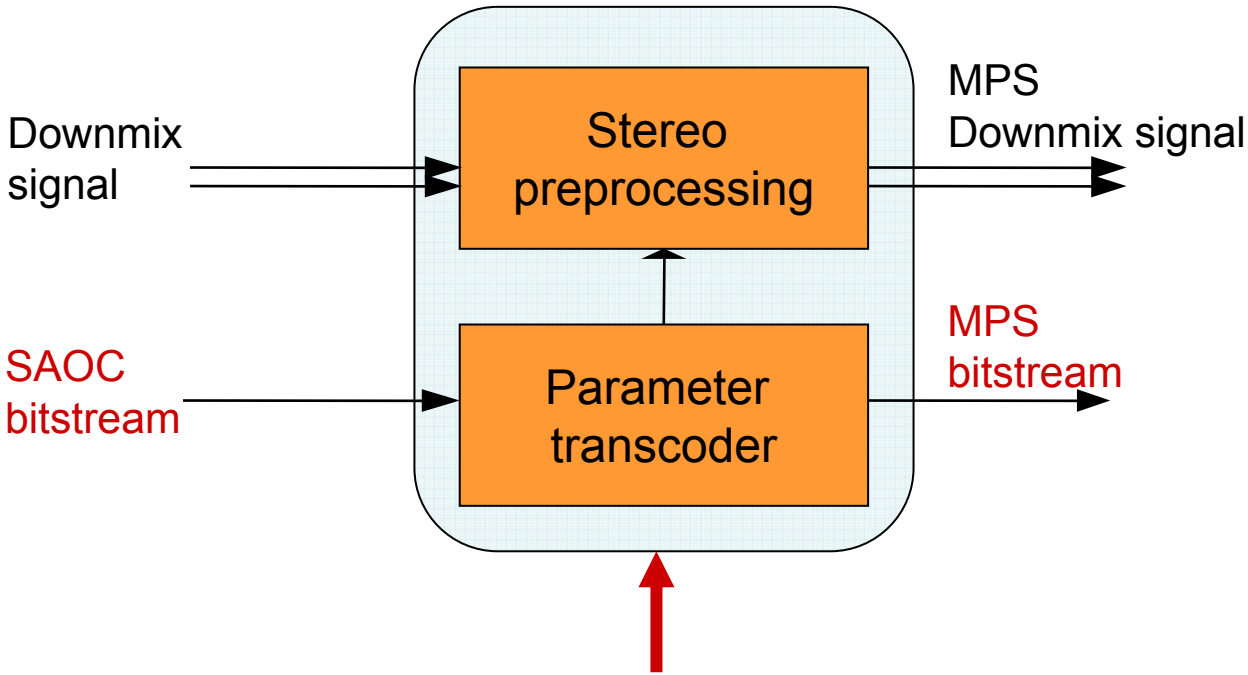


coding technologies

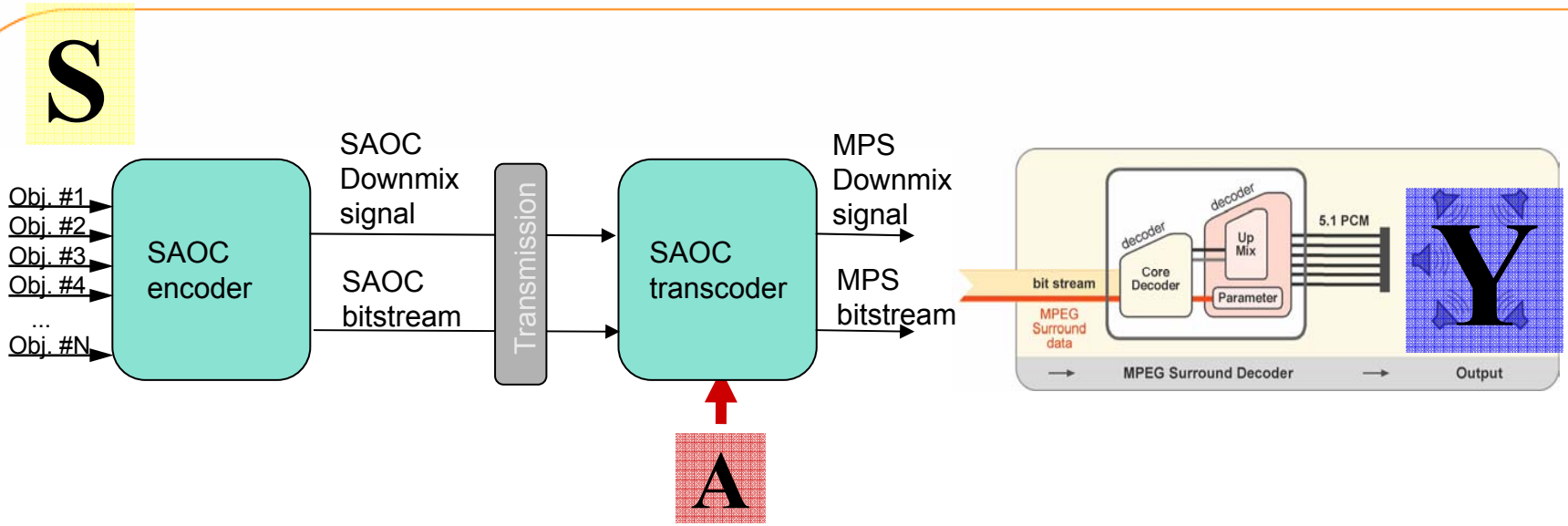
Stereo downmix mode

- Better quality (object separation)
- Downmix is created in encoder
- Allows full backward compatibility to stereo systems

SAOC transcoder: Stereo downmix



Technical enhancements – Stereo downmix



Rendering matrix

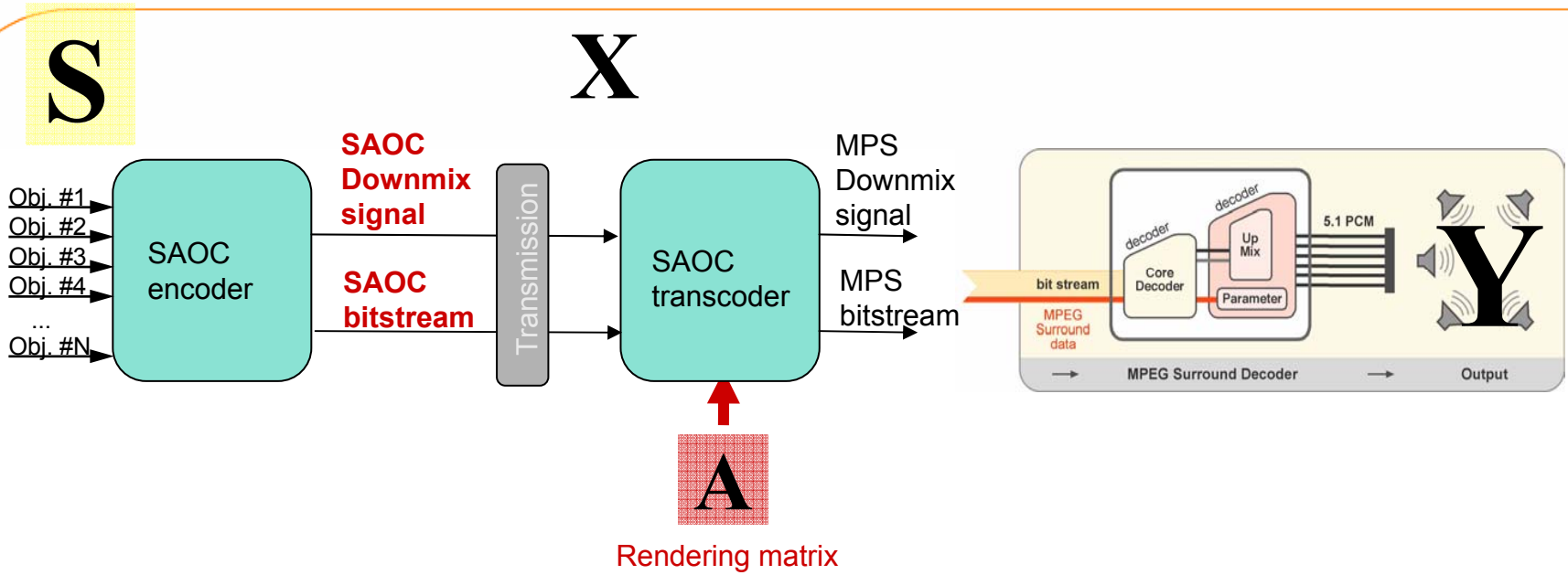
N audio objects:
(1 time-frequency tile)

$$\mathbf{S} = \begin{pmatrix} s_1 \\ \vdots \\ s_N \end{pmatrix}$$

Target rendering:

$$\mathbf{Y} = \mathbf{AS} = \begin{pmatrix} a_{11} & \cdots & a_{1N} \\ a_{21} & \cdots & a_{2N} \\ a_{31} & \cdots & a_{3N} \\ a_{41} & \cdots & a_{4N} \\ a_{51} & \cdots & a_{5N} \end{pmatrix} \begin{pmatrix} s_1 \\ \vdots \\ s_N \end{pmatrix} = \begin{pmatrix} lf \\ ls \\ rf \\ rs \\ c \end{pmatrix}$$

Technical enhancements – Stereo downmix



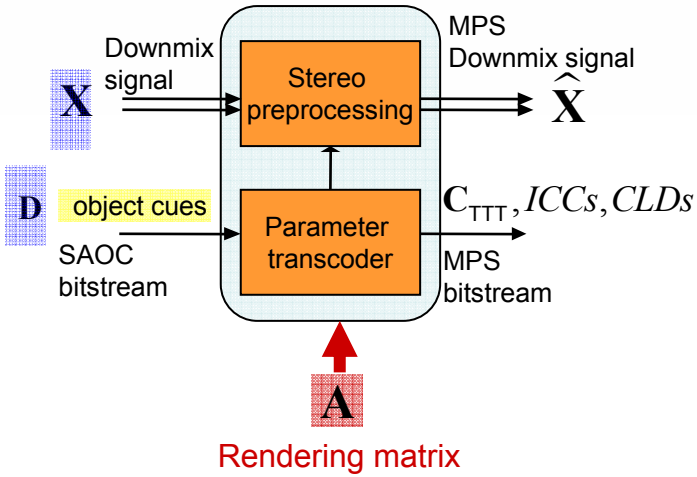
SAOC bitstream:

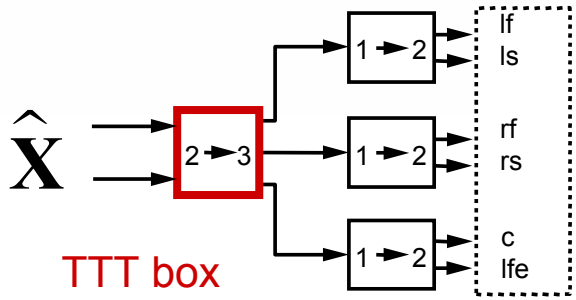
- Object cues:
  - Object energies
  - Inter-object correlations
- Downmix matrix $\mathbf{D}$

SAOC downmix:

$$\mathbf{X} = \mathbf{DS} = \begin{pmatrix} d_{11} & \cdots & d_{1N} \\ d_{21} & \cdots & d_{21} \end{pmatrix} \begin{pmatrix} s_1 \\ \vdots \\ s_N \end{pmatrix} = \begin{pmatrix} l_{dmx} \\ r_{dmx} \end{pmatrix}$$

Technical enhancements - Stereo downmix - TTT





Step 1:

Find 3x2 upmix matrix, $\mathbf{C}_3$

$$\mathbf{A}_3\mathbf{S} \approx \mathbf{C}_3\mathbf{X}$$

$$\mathbf{X} = \mathbf{D}\mathbf{S} \qquad \mathbf{S}\mathbf{S}^* \approx \mathbf{E}$$

$$\mathbf{C}_3 = \mathbf{A}_3\mathbf{E}\mathbf{D}^*\left(\mathbf{D}\mathbf{E}\mathbf{D}^*\right)^{-1}$$

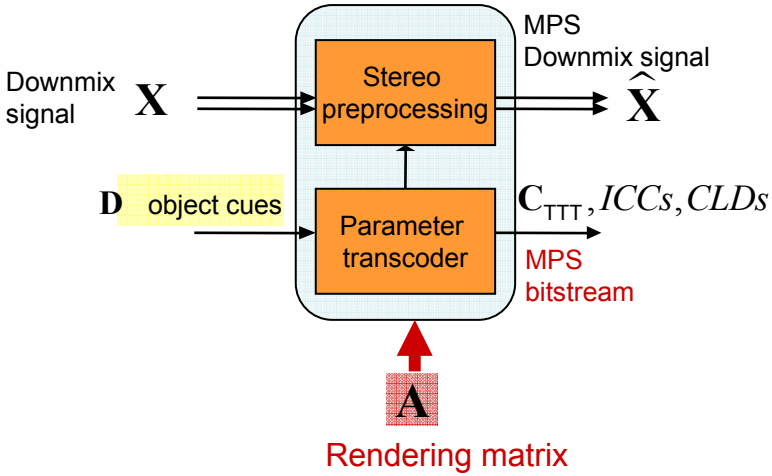
Step 2:

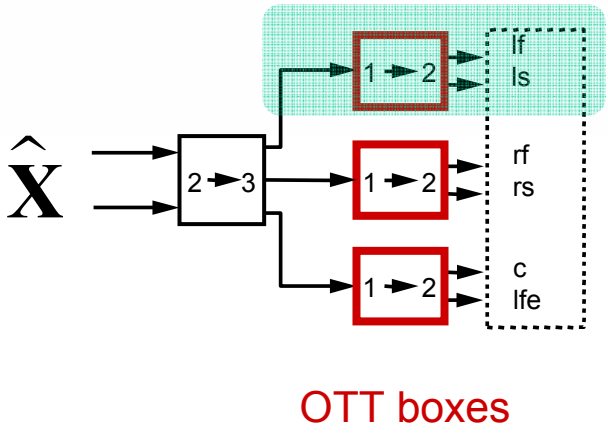
Factorization into MPS upmix matrix $\mathbf{C}_{TTT}$ and stereo preprocessing matrix $\mathbf{G}$

$$\mathbf{C}_{TTT}\mathbf{G} = \mathbf{C}_3$$

Preprocessing gives stereo downmix of target rendering,

$$\hat{\mathbf{X}} = \mathbf{G}\mathbf{X}$$





Step 3:

Parameters for OTT boxes from target covariance matrix $\mathbf{F}$

$$\mathbf{F} = \mathbf{YY}^* =$$

$$= \mathbf{A}(\mathbf{SS}^*)\mathbf{A}^* = \mathbf{AEA}^*$$

$$\mathbf{SS}^* \approx \mathbf{E}$$

$$\mathbf{F} = \begin{array}{c} \\ lf \\ ls \\ \\ \\ \\ \end{array} \begin{array}{c} \begin{array}{cccccc} lf & ls & & & & \end{array} \\ \begin{pmatrix} f_{11} & f_{12} & \cdots & \cdots & f_{15} \\ f_{21} & f_{22} & & & \vdots \\ \vdots & & \ddots & & \vdots \\ & & & \ddots & \vdots \\ f_{51} & \cdots & \cdots & \cdots & f_{55} \end{pmatrix} \end{array}$$

$$CLD_{1,2} = 10\log\left(\frac{f_{1,1}}{f_{2,2}}\right) \qquad ICC_{1,2} = \frac{(f_{1,2})}{\sqrt{f_{1,1}f_{2,2}}}$$

Summary

- Applications: Teleconferencing, gaming, re-mix
- Transcoding enables efficient re-use of MPEG Surround
- Stereo downmix mode allows for
  - flexible downmix schemes
  - stereo backward compatibility
  - better quality

Thank you for your attention!