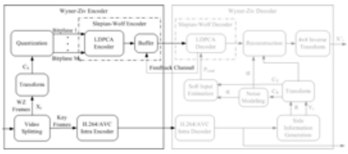
## Distributed Video Coding (DVC) Perspectives

1. Video Coding in Communication Networks
   - simple encoders
   - scalability (SVC)
   - distributed (DVC)
   - error tolerance



2. New applications
   - stereo, 3D, multiview, camera-array TV/video
   - sensor networks

*DVC Encoder, Transform Domain Wyner-Ziv (Stanford Architecture):*

Key frames: H.264 AVC

WZ frames in between: 4x4 transform, bit-planes, LDPCA



## Performance

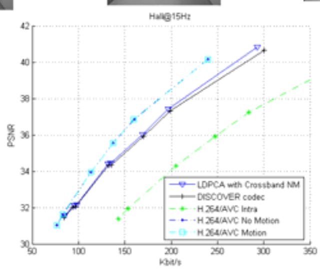
DVC is a promising video coding solution for resource critical applications, e.g encoding without motion compensation.

Visual comparison between no-motion Inter, WZ and Intra coded frames

| H.264/AVC No-motion Inter | Our DVC codec | H.264/AVC Intra |
|---|---|---|
| 82.05 Kbit/s | 80.53 Kbit/s | 78.16 Kbit/s |
| 41.69 dB | 40.85 dB | 31.95 dB |

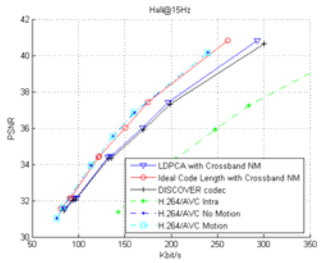


## Performance – low motion





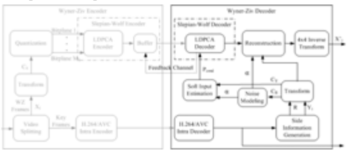

## Performance – low motion




Gap may (to some extent) be explained by LDPCA performance, e.g. finite block length, incremental redundancy, …

## Decoding Procedure (cont.)

5. LDPCA decoder
   - Sum-Product Algorithm
   - Feedback channel
   - Convergence:
     - 8 bits CRC




     - Received syndrome vs. the syndrome of decoded bitplane.

6. Reconstruction
   - Expectation of $C_x$ given the decoded quantization index and side information value, $C_y$.

$$E(C_X \mid [q_i Q_{step}, q_{i+1} Q_{step}), C_Y)$$

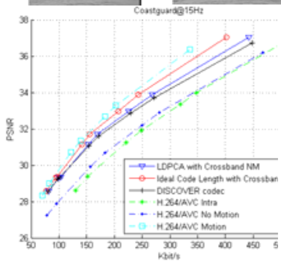
   - Reconstruction is performed using closed form expressions derived from Laplacian distribution model

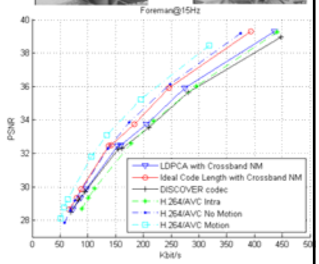
$$C_X = \frac{\int_{q_i Q_{step}}^{q_{i+1} Q_{step}} z f(z) dz}{\int_{q_i Q_{step}}^{q_{i+1} Q_{step}} f(z) dz}$$

7. Inverse Transform



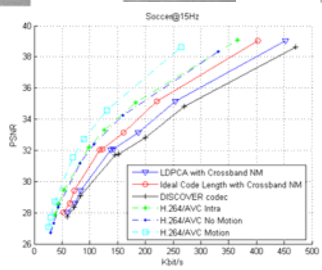
## Performance – medium motion






## Performance – high motion






## Noise (Residual) Models

- Laplacian Distribution
  - Motion estimated residue $R_{ME}$ $\quad f(X_{2i}(x,y) - Y_{2i}(x,y)) = \frac{\alpha_0}{2} e^{-\alpha_0 |X_{2i}(x,y) - Y_{2i}(x,y)|} \approx \frac{\alpha}{2} e^{-\alpha |R_{ME}(x,y)|}$
  - Laplacian parameter
  - Take temporal and spatial variations into account
- Band Level $\quad \alpha_{b_k}^{|\sigma|} = \sqrt{2/\sigma_{|b_k|}^2}, \quad \sigma_{|b_k|}^2 = E(|C_{R_{ME}}^{b_k}|^2) - E(|C_{R_{ME}}^{b_k}|)^2$
- Coefficient Level (DISCOVER)
  - Evaluate variance of $R_{ME}$
  - Two categories

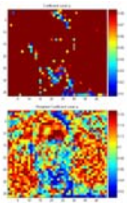
$$\alpha_{b_k}^{c1}(u,v) = \begin{cases} \alpha_{b_k}^{|\sigma|} & if \quad (u,v) \in map_{b_k}^{in} \\ \sqrt{2/D(u,v)^2} & if \quad (u,v) \in map_{b_k}^{out} \end{cases}$$

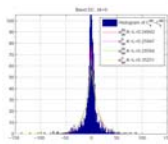
*where*

$$map_{b_k}^{in} = \{(u,v) \mid D(u,v)^2 \le \sigma_{|b_k|}^2\}$$

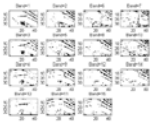
$$map_{b_k}^{out} = \{(u,v) \mid D(u,v)^2 > \sigma_{|b_k|}^2\}$$

$$D(u,v) = C_{R_{ME}}^{b_k}(u,v) - E(|C_{R_{ME}}^{b_k}|)$$





## Improved Noise Model

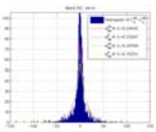


- Refined classification map
  - $R_{ME}$ is unreliable for classification
  - Cross-band correlation
  - Refined noise residue ($R_{ZY}$) between partly decoded frame ($Z$) and side information frame ($Y$) is closer to $R_{XY}$ than $R_{ME}$ within decoded band.
  - A classification map is obtained by comparing the refined values of $D$ and variance by using the $R_{ZY}$

$$\sigma^2_{|b_k|} = E(|C^{b_k}_{R_{ZY}}|^2) - E(|C^{b_k}_{R_{ZY}}|)^2 \qquad map^{in}_{b_k} = \{(u,v) \mid D(u,v)^2 \le \sigma^2_{|b_k|}\}$$

$$D(u,v) = C^{b_k}_{R_{ZY}}(u,v) - E(|C^{b_k}_{R_{ZY}}|) \qquad map^{out}_{b_k} = \{(u,v) \mid D(u,v)^2 > \sigma^2_{|b_k|}\}$$

  - Classification map of higher frequency band is predicted by using decoded lower frequency band map.



- Estimators
  - Variance based estimator
  - The maximum likelihood estimator

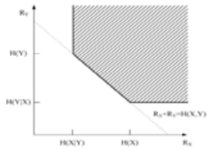
$$\alpha^{|b|}_{b_k} = ((\sum || C^{b_k}_{R_{ME}} | - E(|C^{b_k}_{R_{ME}}|)|)/N)^{-1}$$



## Slepian-Wolf Theorem

Two correlated source $(X, Y)$ ~ i.i.d

- Independent encoding and decoding
  - $R_X \ge H(X)$, $R_Y \ge H(Y)$
- Joint encoding and decoding
  - $R = R_X + R_Y \ge H(X, Y)$
- Separate encoding + joint decoding





$$R_X \ge H(X|Y)$$
$$R_Y \ge H(Y|X)$$
$$R_X + R_Y \ge H(X,Y)$$

Wyner-Ziv extended S-W to lossy coding

Separate encoding + joint decoding is as efficient as joint encoding + decoding!



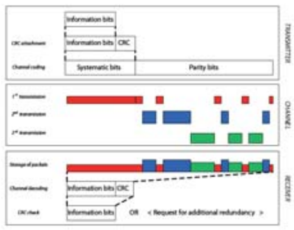
## Error-Correction in Communication

- LDPCA for distributed source/video coding
- 100 Gbit/s forward error correction for optical communication
- Turbo decoding for ESA space projects
- Turbo decoding in LTE – data compression of soft information in HARQ



## Maximum Mutual Information Vector Quantization of Log-likelihood Ratios for Memory Efficient Hybrid ARQ Implementations

- Hybrid ARQ
  - Advanced error-control mechanism used, among others, in 3GPP Long Term Evolution
- Principle
  - A selection of bits of the codeword is sent, more follow if the decoding fails
  - Previously received bits are not discarded, but stored and combined to new information for increased correct decoding probability

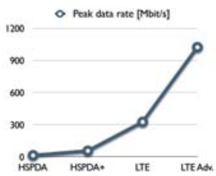



M. Danieli e.a. DCC 2010

## Motivation

- Memory issue
  - Soft-decoding is used, hence the receiver associates a confidence to each bit in the form of Log-Likelihood Ratios
  - Long Term Evolution comes with a sharp increase in data rates
- The memory needed to implement HARQ techniques is bound to grow significantly

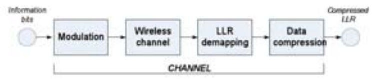
$$l_k = \log \frac{p(x_k = 1|r)}{p(x_k = 0|r)}$$






## Problem statement

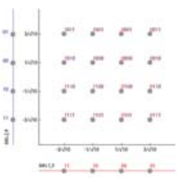
- Find a suitable *data compression algorithm* for log-likelihood ratios in order to *decrease memory requirement* while limiting the increase in computational complexity and the performance loss
- The data compression algorithm becomes part of a *channel* whose capacity it makes sense to maximize

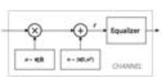





## System model

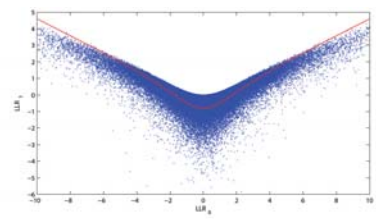
- Reference technology
  - 3GPP Long Term Evolution
- Our model is based on the LTE standard
  - *Modulation:* QPSK, 16-QAM and 64-QAM that correspond to BPSK, 4-PAM and 8-PAM when projected on the real axis
  - *Channel:* Rayleigh distributed gain, AWGN, perfect equalization
  - *Receiver:* LLR computation








## Log-likelihood ratios



- When the channel is Rayleigh fading, a coupling appears between LLRs of bits mapped on the same symbol
- Visualizing the data (here in the case of 4-PAM modulation) we needed to compress helped finding a suitable algorithm, namely *vector quantization*

## Maximum mutual information VQ

The mutual information is I(X;Y) = H(X) – H(X|Y)

- The capacity is defined as the mutual information between input and output, in this case between the bit $x_k$ and the quantized version of the corresponding LLR, $y_k$
- To maximize the channel capacity we need to maximize the mutual information
- Equivalently, we can choose to minimize the *mutual information loss* incurred when we substitute $y_k$ (the quantizer output) for $l_k$ (the original LLR)



## Kullback-Leibler divergence as cost

- If we use training vectors **t**={$t_1,\ldots,t_K$} to represent the input, we can express the *total distortion* as a sum of *local distortions* between samples and reconstruction values
- This distortion is the *Kullback-Leibler divergence* between the probability distributions associated to the original LLR and its quantized counterpart

$$\begin{aligned}\Delta I &= H(X_k|I) - H(X_k|\mathbf{T}) \\ &= \sum_{\mathbf{t}\in T}\sum_{x_k\in\{0,1\}} \boxed{p(\mathbf{t})}\, p(x_k|\mathbf{t})\log_2 p(x_k|\mathbf{t}) \\ &\quad - \sum_{i=1}^{M}\sum_{x_k\in\{0,1\}} \boxed{p(i)}\, p(x_k|i)\log_2 p(x_k|i)\end{aligned}$$

$p(\mathbf{t}) \rightarrow 1/n_T$, $p(i) \rightarrow n_i/n_T$

$$p(x_k|i) = \sum_{\mathbf{t}\in R_i}\frac{1}{n_i}p(x_k|t_k) = \boxed{q(x_k|\mathbf{t})}$$

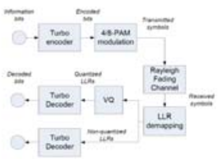
Average posterior in region $R_i$

$$\begin{aligned}\Delta I &= \frac{1}{n_T}\sum_{\mathbf{t}\in T}\sum_{x_k\in\{0,1\}} p(x_k|\mathbf{t})\log_2\frac{p(x_k|\mathbf{t})}{q(x_k|\mathbf{t})} \\ &= \frac{1}{n_T}\sum_{\mathbf{t}\in T} D_{\mathrm{KL}}(p_{x_k|\mathbf{t}}\,||\,q_{x_k|\mathbf{t}})\end{aligned}$$



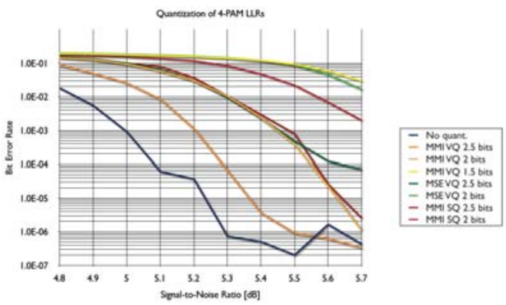
## Simulation set-up

- The codebook design procedure guarantees by construction to reach at least a local maximum for the capacity
- We need to verify that increased capacity leads to improved performance
  - We insert error-correcting coding in our model
  - In particular, we use the rate 1/3 turbo code defined in the LTE standard (constituent encoder 15/13) with block length $l_{bl}$=6144






## Results



## Vision: Real-time global free-view 3D TV



## More info

- DTU Fotonik: www.fotonik.dtu.dk
- Søren Forchhammer sofo@fotonik.dtu.dk
- Studies: http://www.fotonik.dtu.dk/English.aspx
  - PhD call: High quality video over broadband optical-wireless network connections, deadline May 3rd,
  - Summer University
  - Master programme