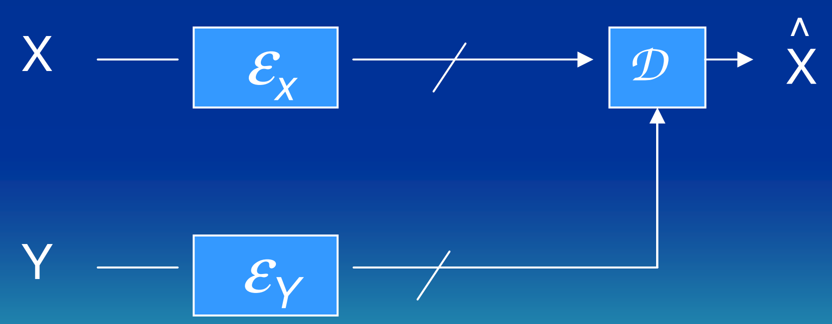
Shared Descriptions Fusion Coding for Storage and Selective Retrieval of Correlated Sources

Sharadh Ramswamy and Kenneth Rose
Signal Compression Lab, UCSB

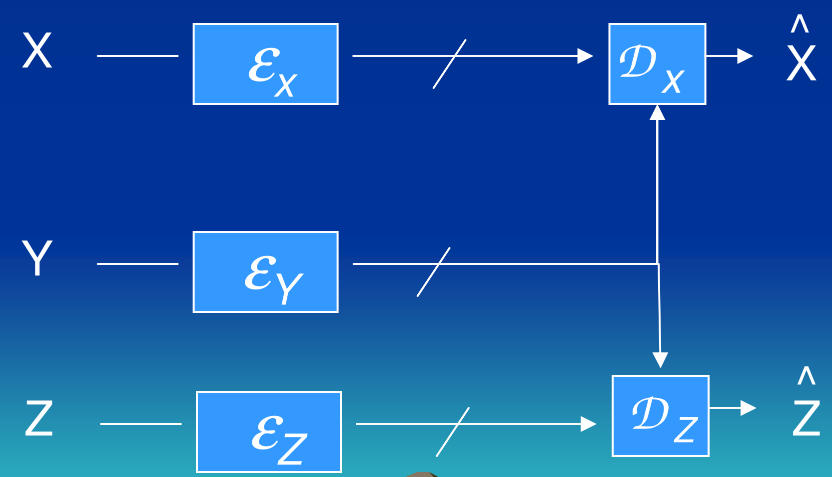
Coding of Correlated Sources

- Well studied problem in Information Theory (Slepian-Wolf (1973), Wyner-Ziv (1976))
- Independent encoding/transmission
- (Joint) Decoding with Side-information



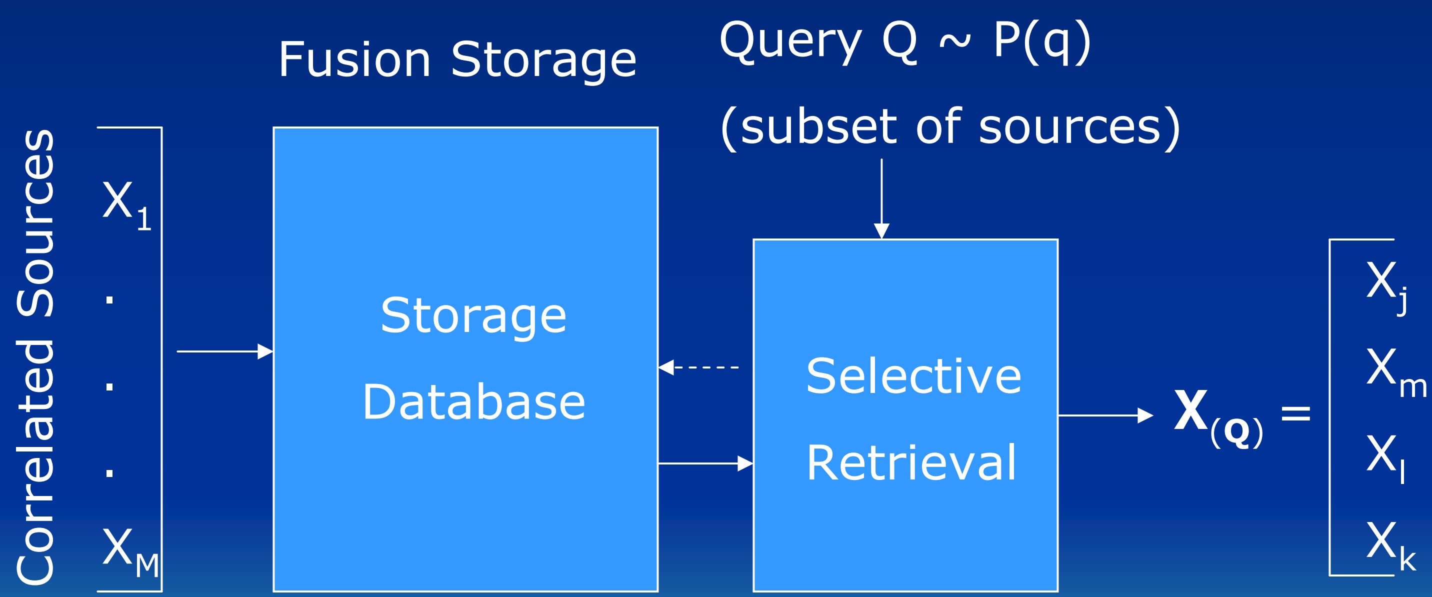
Coding of Correlated Sources …

- Other flavors: multi-terminal source coding, distributed source coding
- Applications: distributed compression in sensor networks (DISCUS (2000), Network VQ (2001))



Coding Correlated Sources for Storage

- New setting: *Storage* Media
- **Joint** encoding/compression/storage of sources
- **Selective** Retrieval of sources!!!



Storage Rate R_s vs. Retrieval Rate R_r

Min. Storage Rate vs. Min. Retrieval Rate

- Compressing all sources together minimizes storage

$$R_s=H(X_1,\ldots,X_M)$$

  - but compromises retrieval speed

$$R_r=R_s=H(X_1,\ldots,X_M)>> \Sigma_q P(q)H(X_{(q)})$$

- Compressing each subset separately minimizes retrieval rate/time

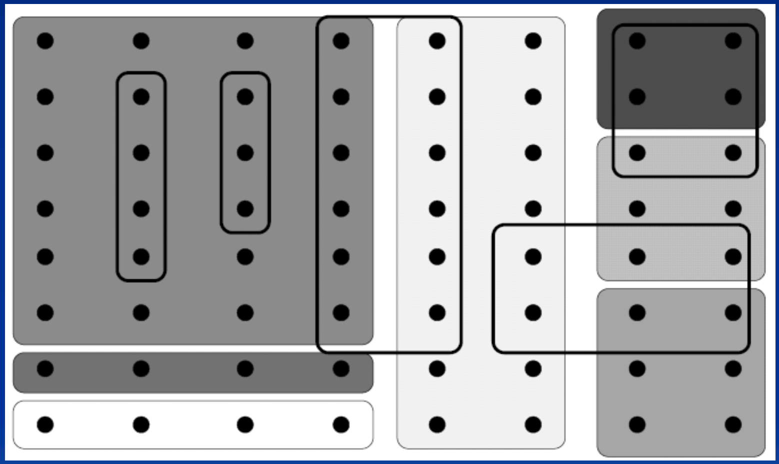
$$R_r= \Sigma_q P(q)H(X_{(q)})$$

  - but (exponentially) large query sets result in very high storage rate

$$R_s= \Sigma_q H(X_{(q)})>>H(X_1,\ldots,X_M)$$

Impact/Applications

- Storage, search and retrieval of correlated streams of data e.g. from sensor networks, stocks



A 2D Sensor Field: boxes are regions of interest

Prior Work on Fusion Storage Coding

- Asymptotically lossless “fusion codes” analyzed by Nayak et. al. (2005)
- Reformulation as a multi-terminal source coding problem (Han and Kobayashi (1980))
- A single letter achievable rate region also given

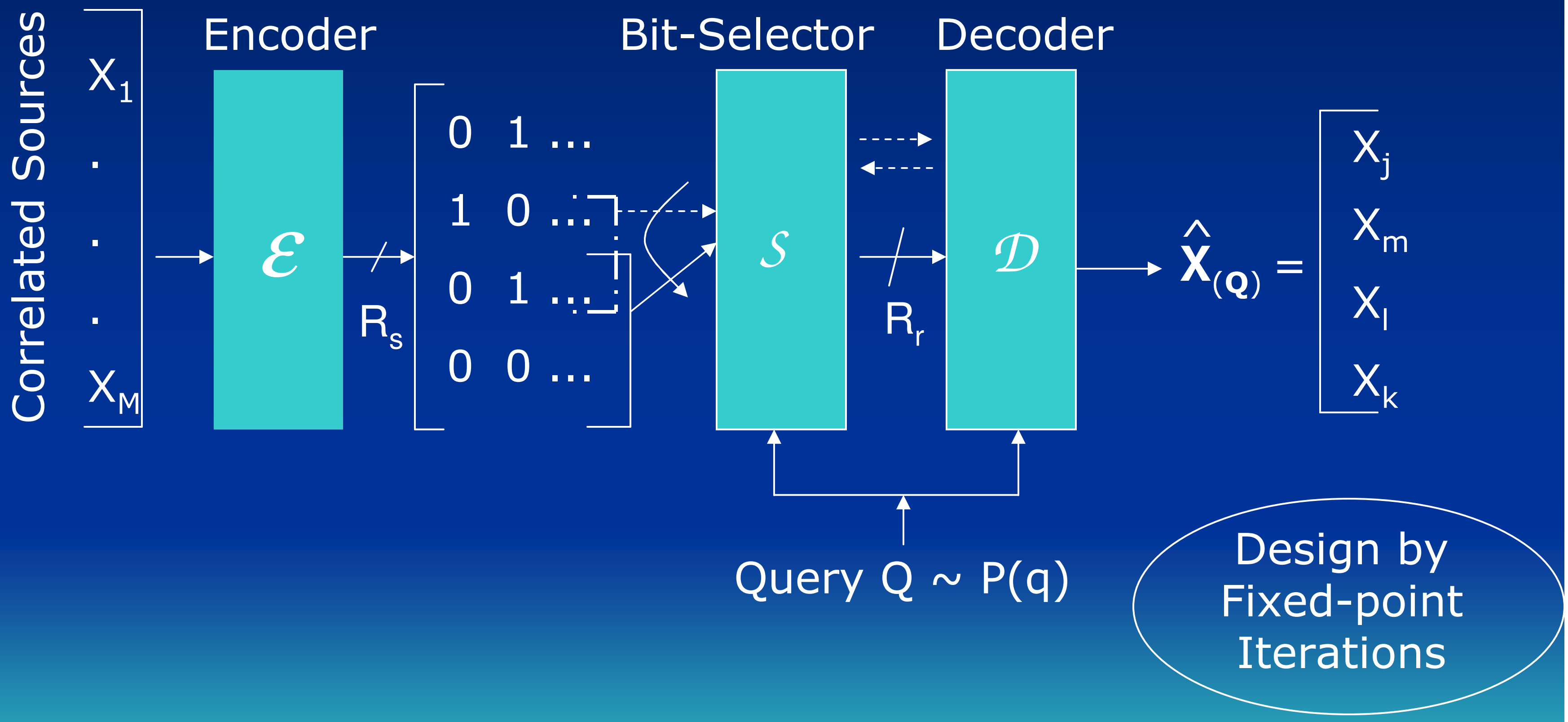
Practical Fusion Coding

- Fusion Coders by Ramaswamy et. al. (2007)
- Storage of signals – with (lossy) quantization
- Storage devices have fixed (limited) storage capacity (R_s)
- Allowed R_s, trade-off between distortion and retrieval rate optimized:

$$\min D(R_s) + \lambda R_r(R_s)$$

- *Query-dependent* bit-(subset) selection (and relevant codebooks) for selective retrieval …

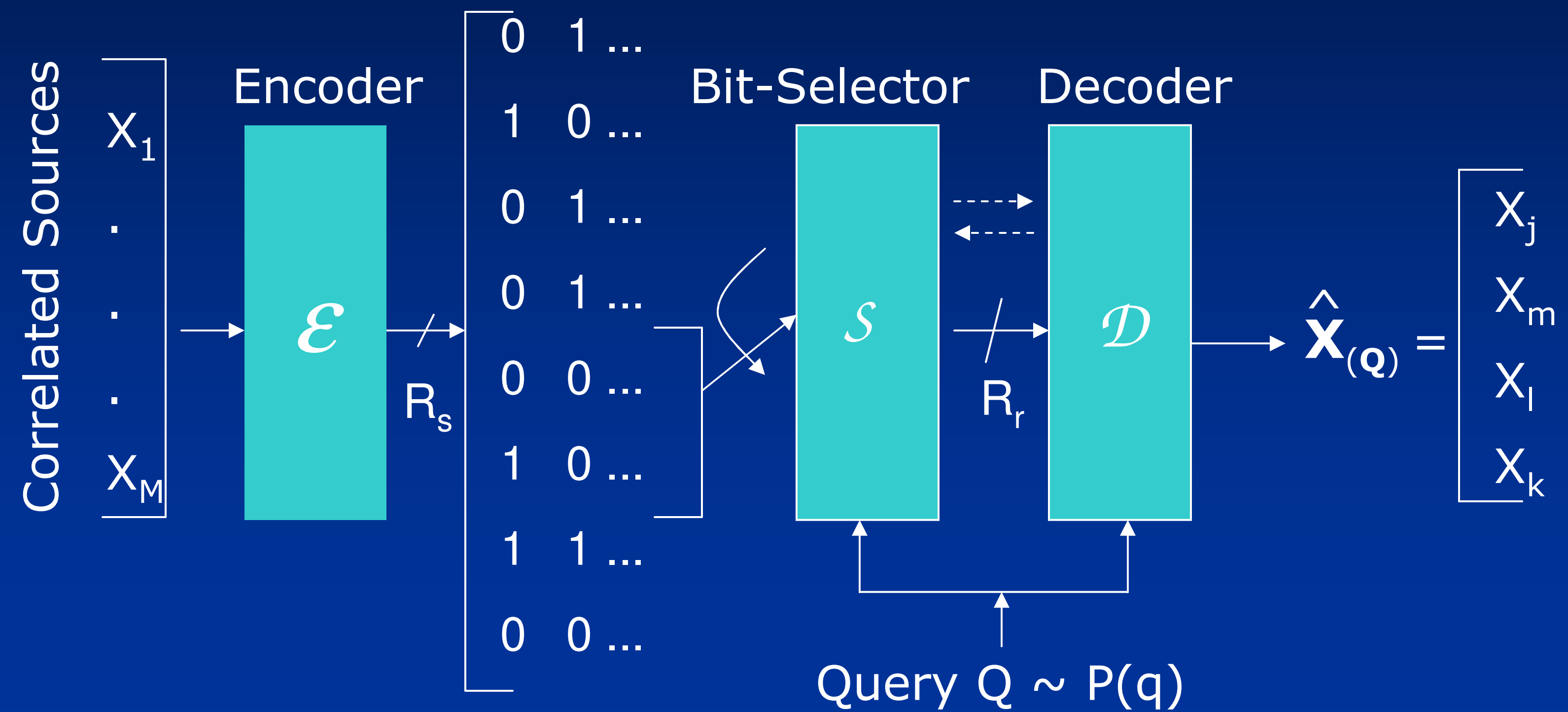
The Fusion Coder (FC)



Advantages and Limitations of FC

- Significant gains over joint comp. (VQ)
- Better performance at higher R_s
  - -needed for large sensor networks
- Higher R_s => more freedom to design bit-selector
- But system complexity ~ $O(2^{R_s})$

Scalability of Fusion Coder



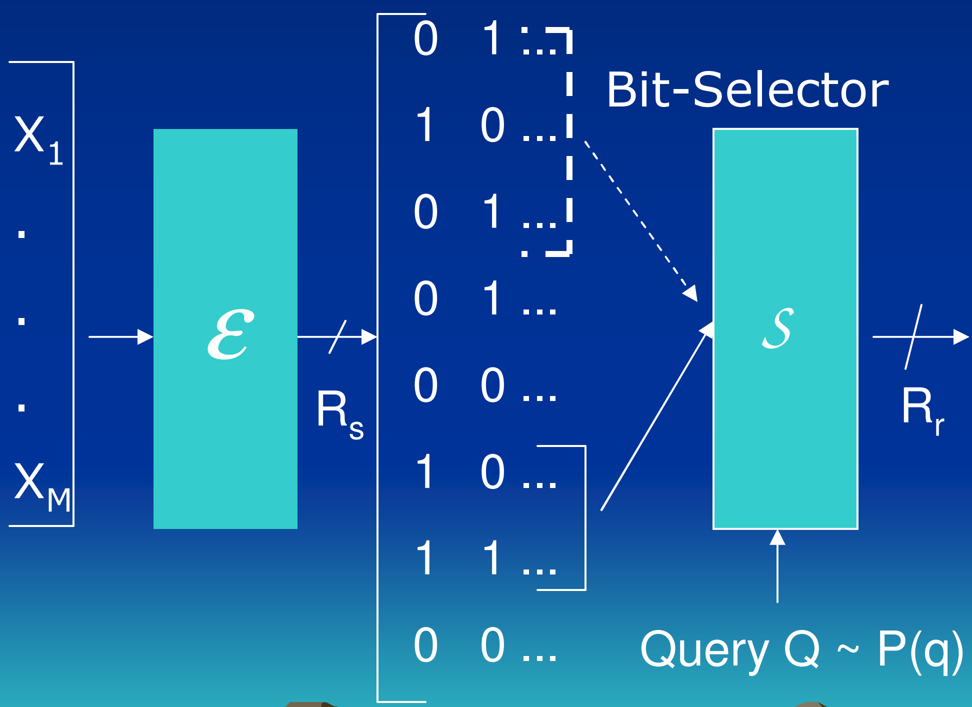
Encoder operation: $O(2^{R_s})$ indices

Codebook storage: $O(2^{R_s})$ codevectors

Bit-selector design: $O(2^{R_s})$ bit-combinations

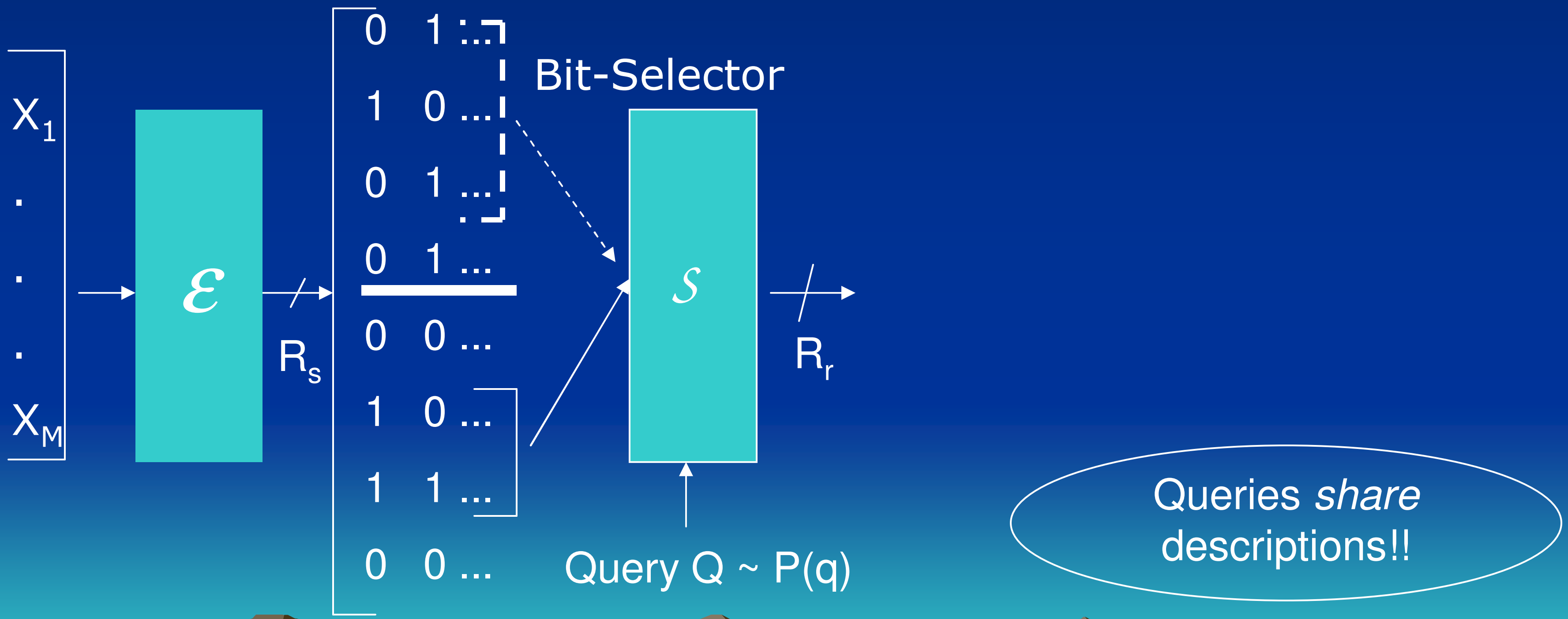
The Shared Descriptions Approach

- Impose structural constraints on bit-selector module
- Selection *only* from disjoint *groups of bits* (descriptions)



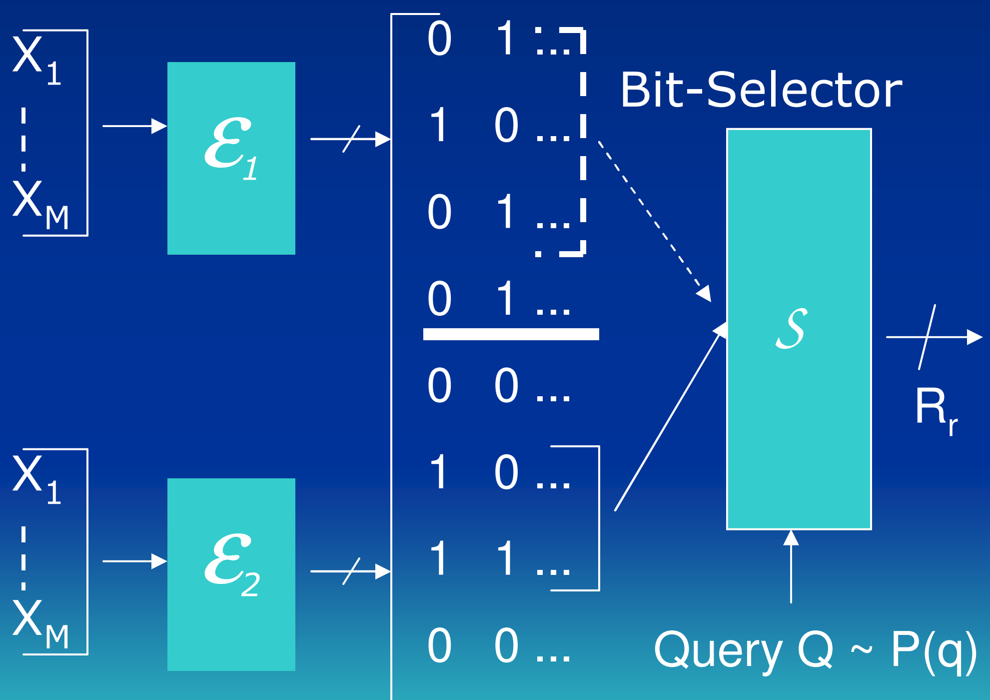
The Shared Descriptions Approach

- Impose structural constraints on bit-selector module
- Selection *only* from disjoint *groups of bits* (descriptions)



The Shared Descriptions Approach

- Impose structural constraints on bit-selector module
- Selection *only* from disjoint *groups of bits* (descriptions)



- Encoding of disjoint groups done independently!!
- Storage requirement smaller!!
- System complexity: $O(2.2^{(R_s/2)})$

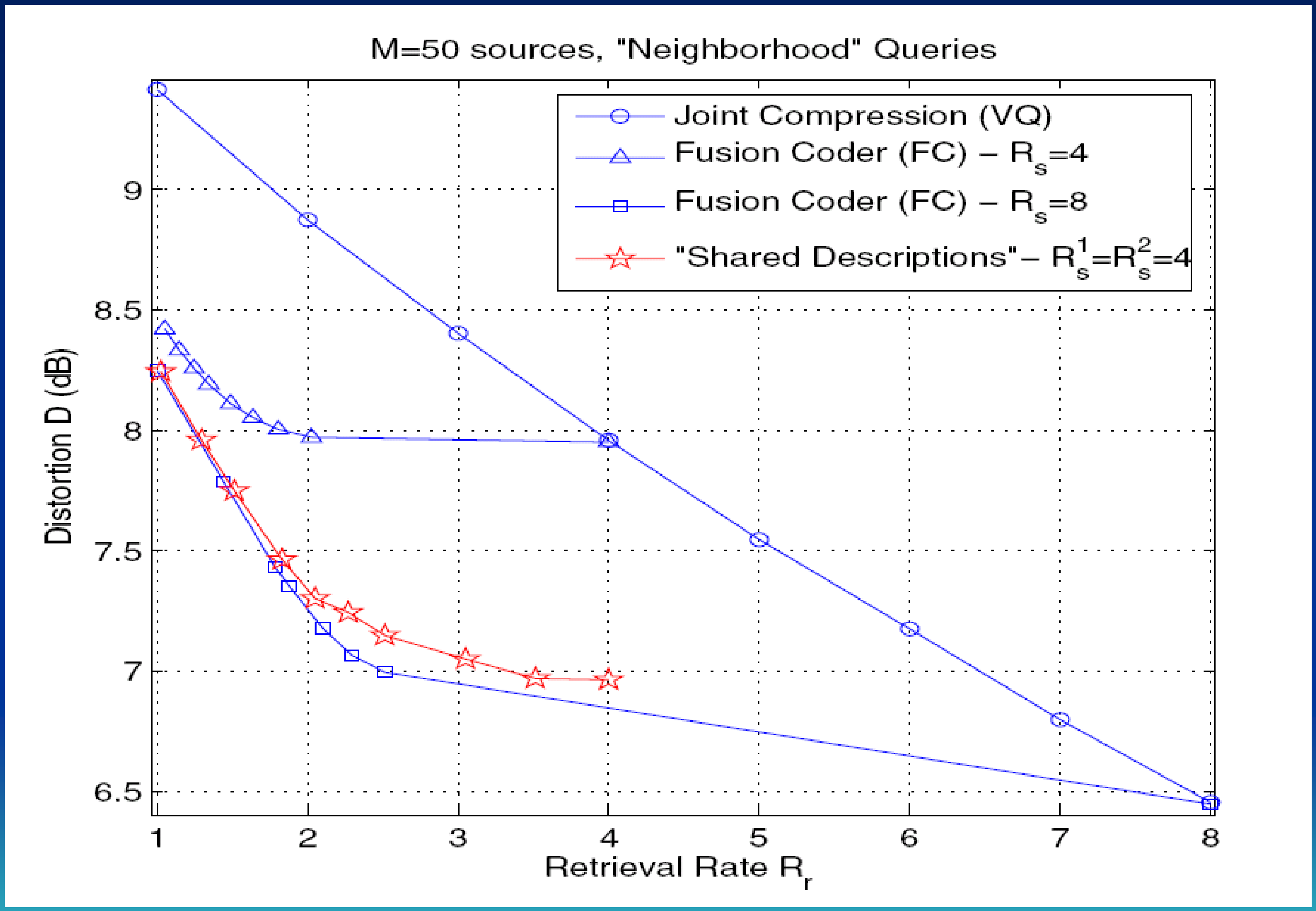
Queries *share* descriptions!!

Experimental Set-up

- Sources: Zero-mean, correlated memoryless Gaussian rv's
- $E(X_iX_j)=\rho_{ij}=\rho^{|i-j|}$ ≡ linear sensor array
- "Neighborhoods" of *n* sources queried

- M=50 sources, ρ=0.8,
- n=10, Uniform query distribution, $|\mathcal{Q}|$=41

Comparison with Fusion Coding



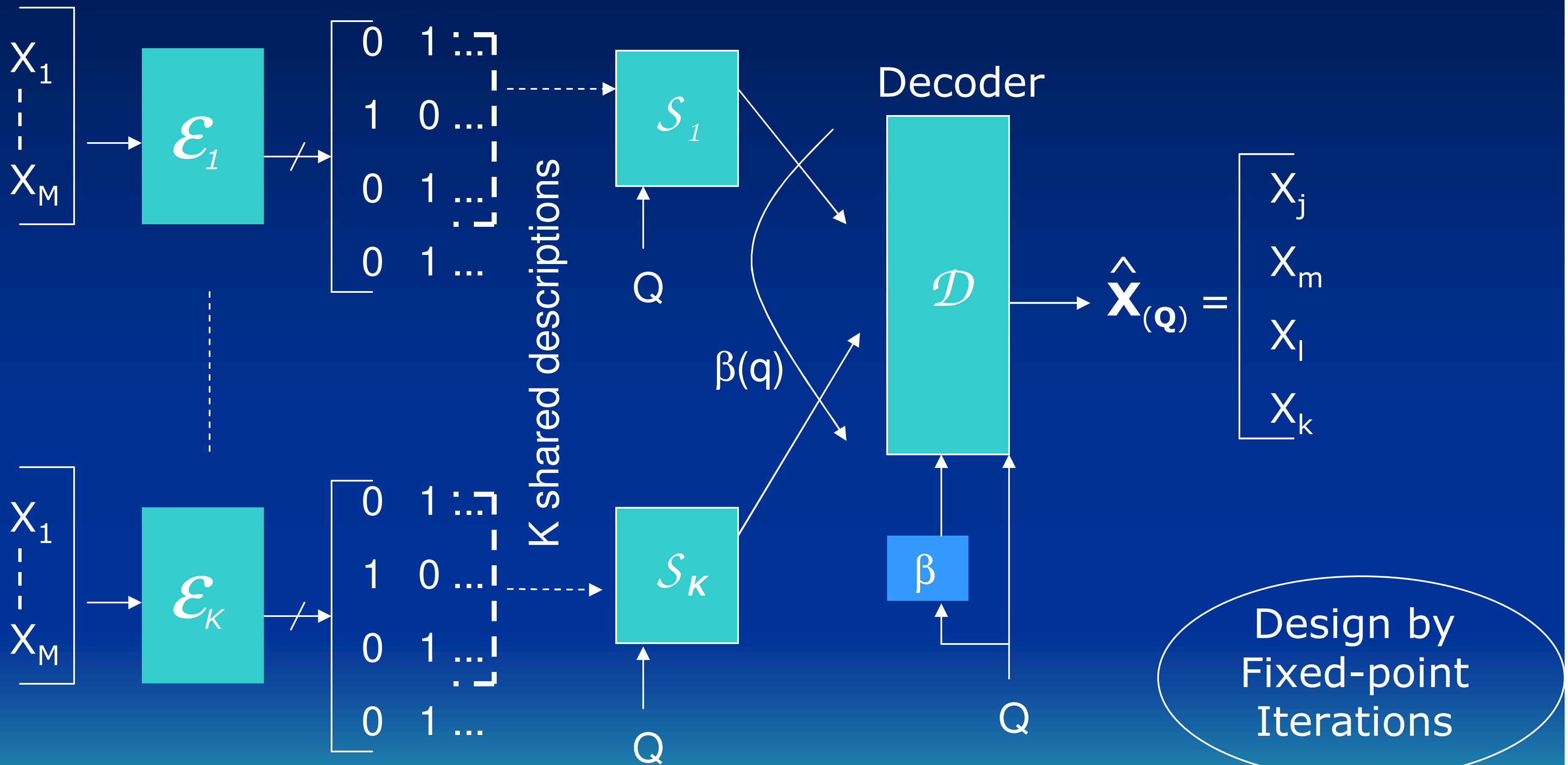
Shared Descriptions Fusion Coding

- Bit-selector = Description selector β

  + (within description) bit-selector $\mathcal{S}_k$
- Bits used by k^{th} description=$R_{s,k}$
- Complexity measure $C_{net} = \Sigma_k 2^{R_{s,k}}$
- Net storage $R_{s,net} = \Sigma_k R_{s,k}$
- Allowed complexity C, storage R_s, K descriptions

$$\min D(R_s) + \lambda R_r(R_s) \ni R_{s,net} \leq R_s,\ C_{net} \leq C$$

Shared Descriptions Fusion Coder

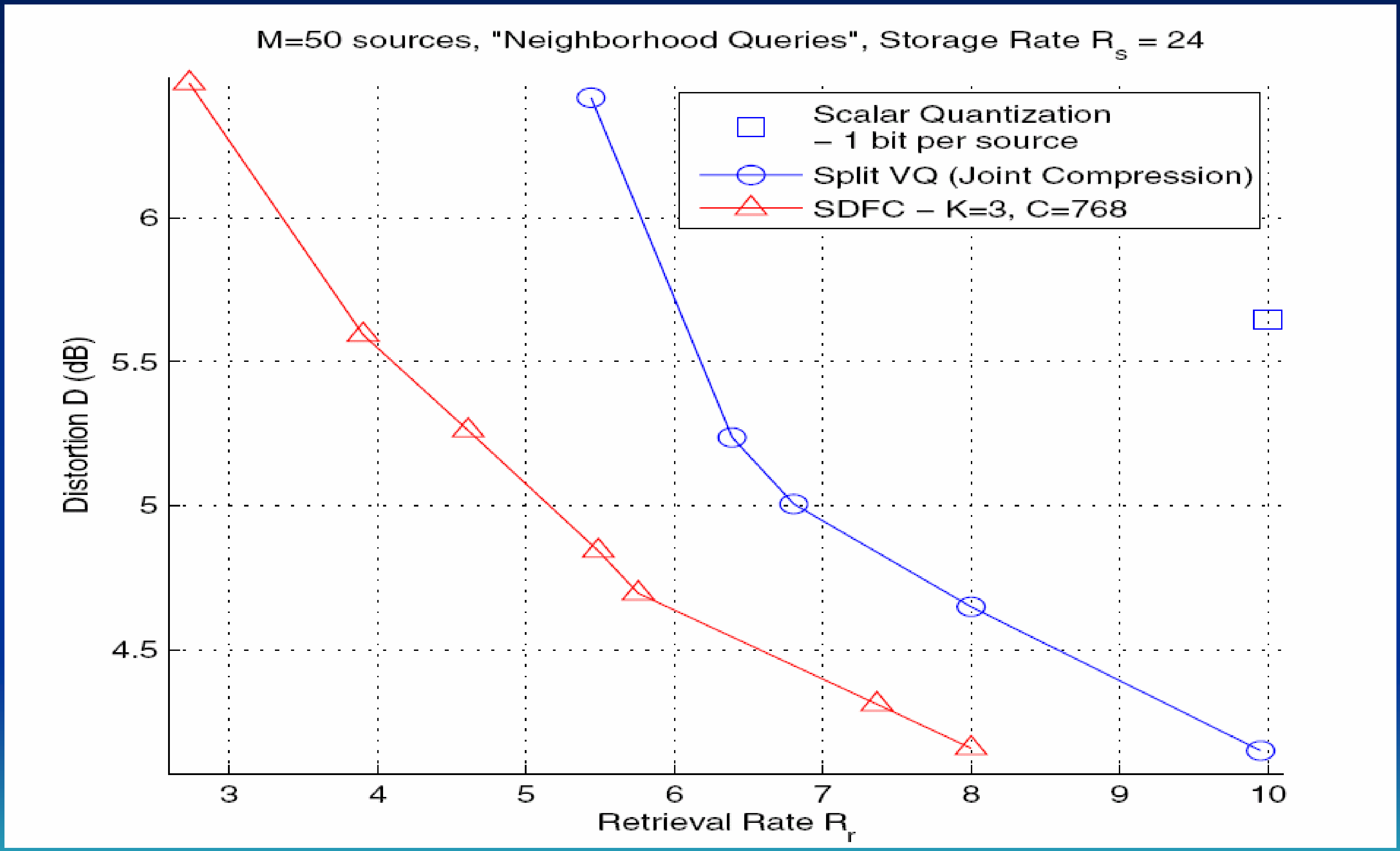


β: Description Selector

Reference for Comparison

- Scalar Quantization:
  - compress each source separately
- Split Vector Quantization (VQ):
  - group sources
  - share/split storage rate
  - compress group
- R_s=24,M=50, Uniform “neighborhood” queries
- Scalar quant. : 1 bit per source
- Split VQ: 24,12,8,6,4 groups

SDFC vs. Split VQ, Scalar Quant.



Conclusions

- Fusion storage and selective retrieval of correlated sources: an important problem
- Fusion coders optimal, but not scalable
- SDFC: “Share descriptions and control complexity”
- Significant advantages over naïve schemes

Future Work

- Evaluate rate-distortion functions
- Quantization in query space
- SDFC + Predictive Fusion Coding …