



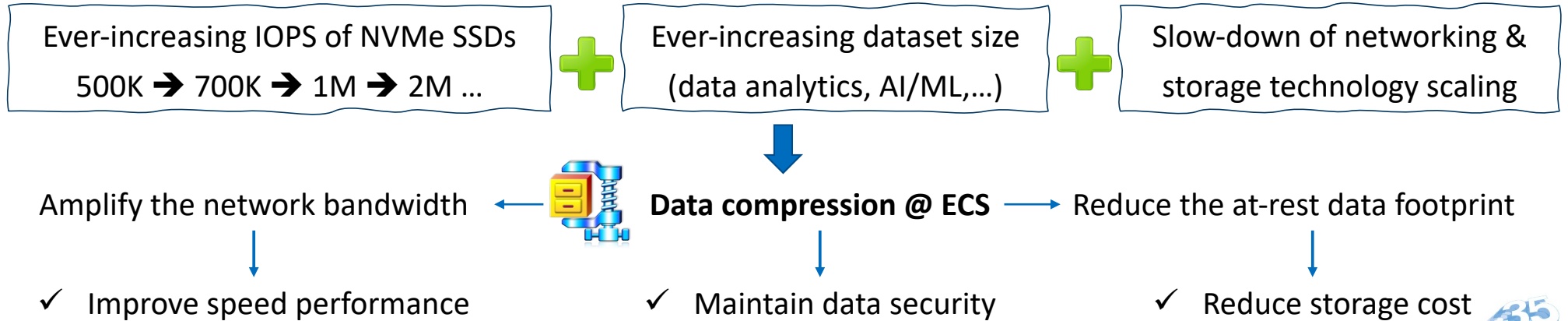
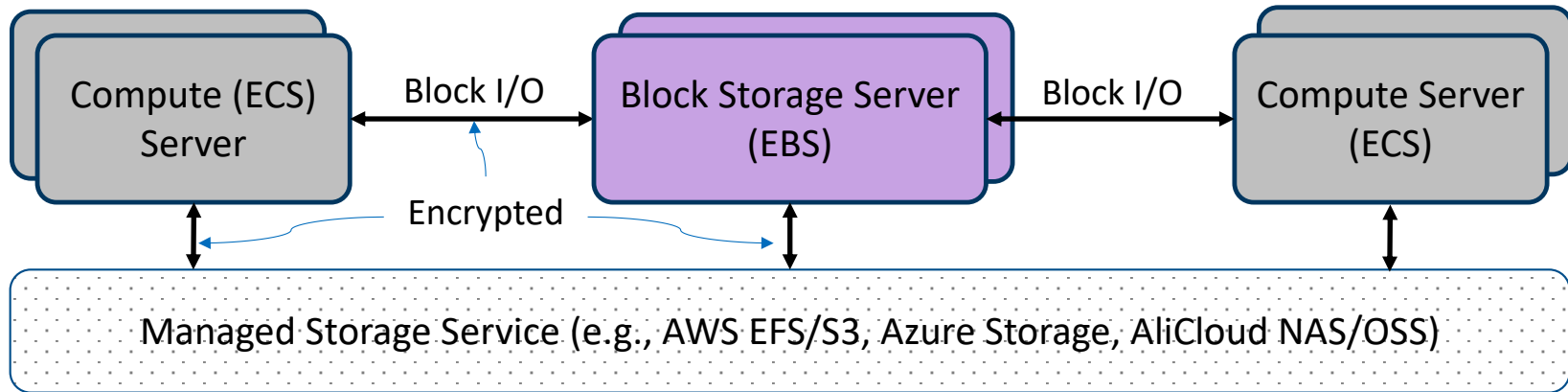
Painlessly Realizing Transparent Compression over Disaggregated Infrastructure

Tong Zhang, ScaleFlux

tong.zhang@scaleflux.com



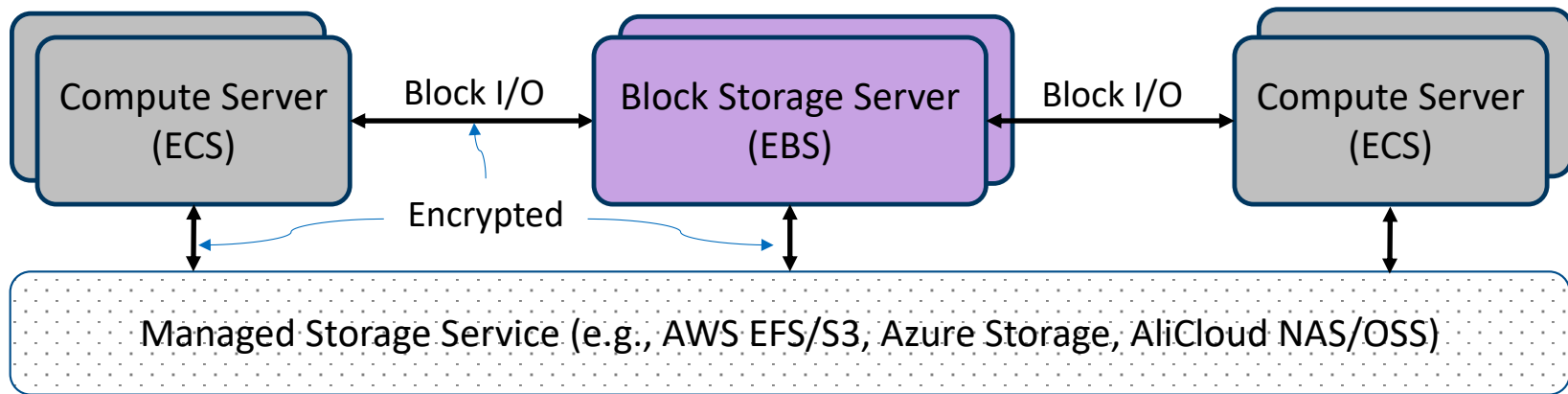
Compute and Storage Disaggregation



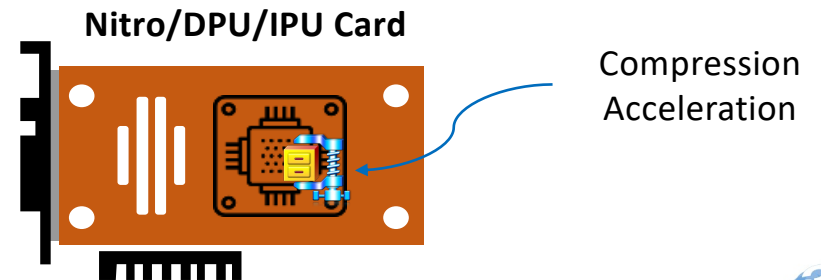
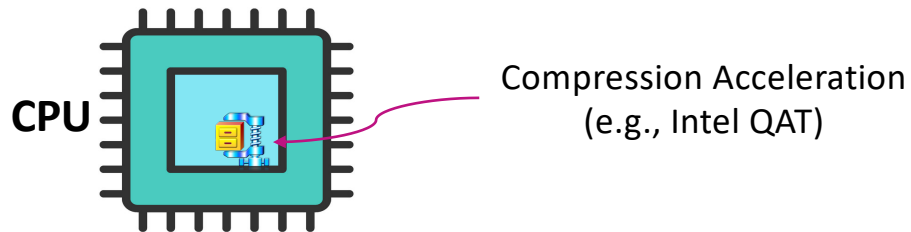
Transparent Compression



Flash Memory Summit



Infrastructure-level compression **transparent** to end users/applications



Transparent Compression

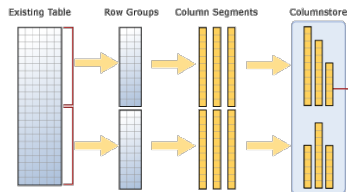


Flash Memory Summit

Difficulty of realizing compression = Difficulty of compressing data + Difficulty of managing compressed data

Data compression granularity \uparrow \rightarrow Difficulty of managing compressed data \downarrow

COLUMN-STORE

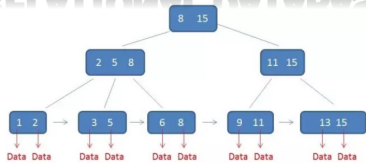


Coarse-grained data access
(e.g., 256KB, 1MB)

Easy management of
compressed data



RELATIONAL DATABASE

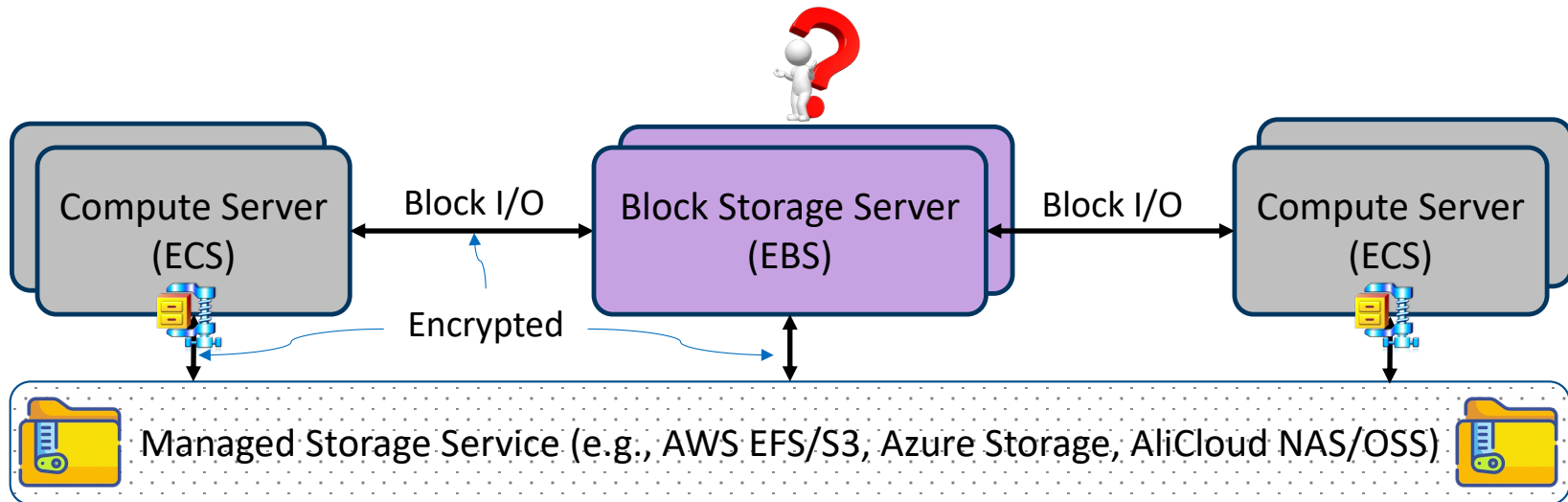


Fine-grained data access
(e.g., 8KB)

Difficult management of
compressed data

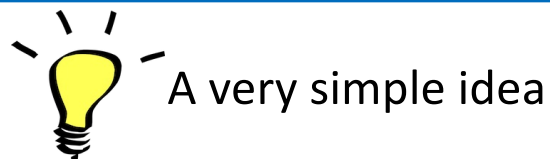


Transparent Compression

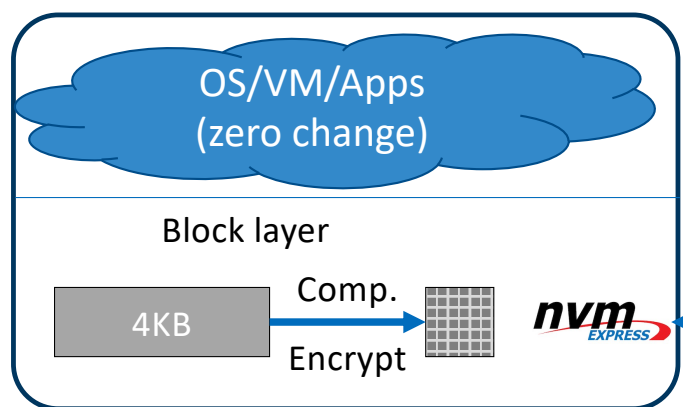


Infrastructure-level compression **transparent** to end users/applications

Disaggregated Transparent Compression

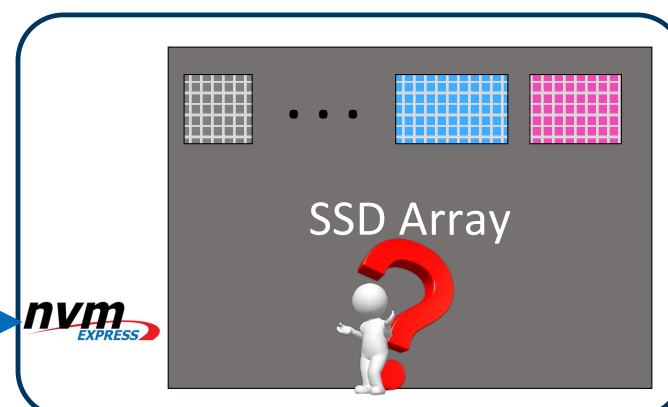


ONLY responsible for comp./encrypting
each 4KB LBA block at the block layer



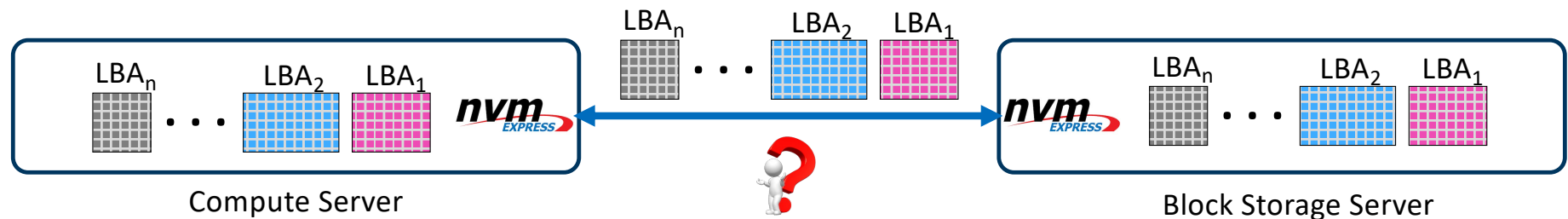
Compute Server

ONLY responsible for managing all the
variable-length data blocks



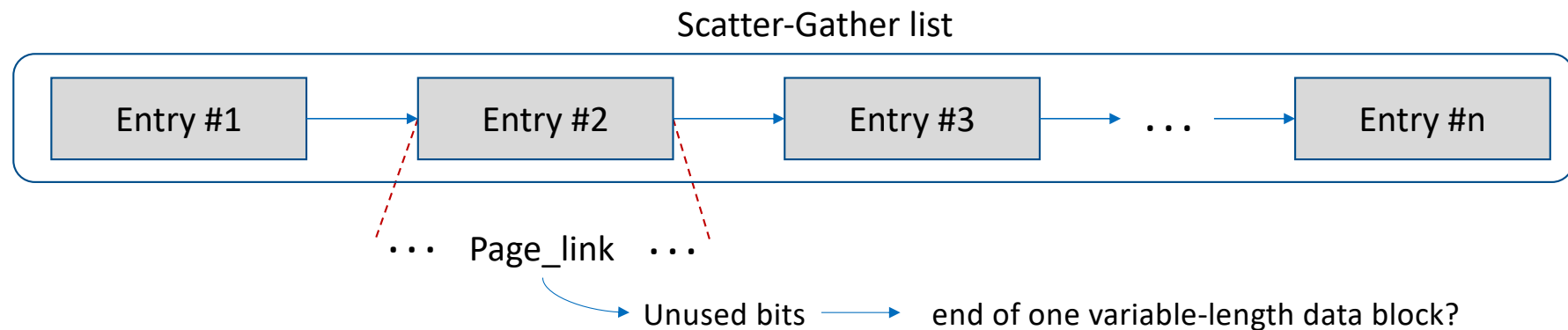
Block Storage Server

NVMe-oF Driver Modification



? Modify the NVMe-oF driver to support variable-length LBA blocks

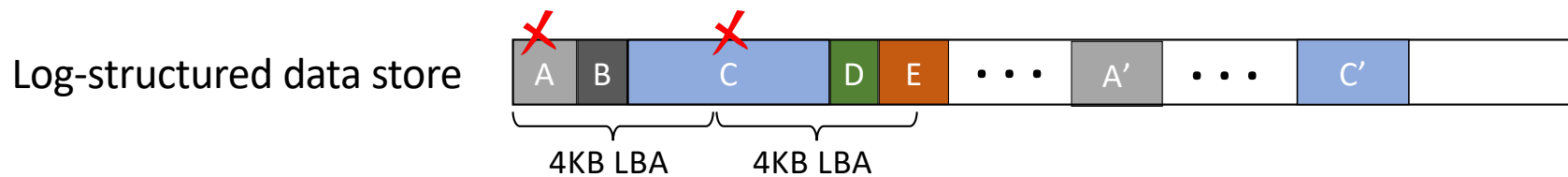
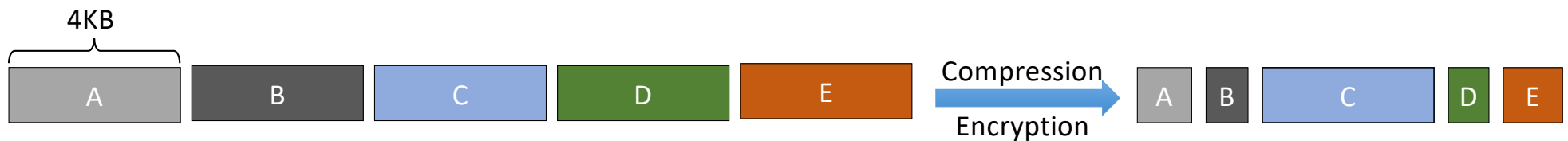
- Must be fully compatible with the API of standard NVMe-oF driver



Block Storage Server Data Management



Storage management of compressed/encrypted variable-length blocks



Garbage collection overhead → IOPS performance & endurance penalty



Block Storage Server Data Management



Storage management of compressed/encrypted variable-length blocks

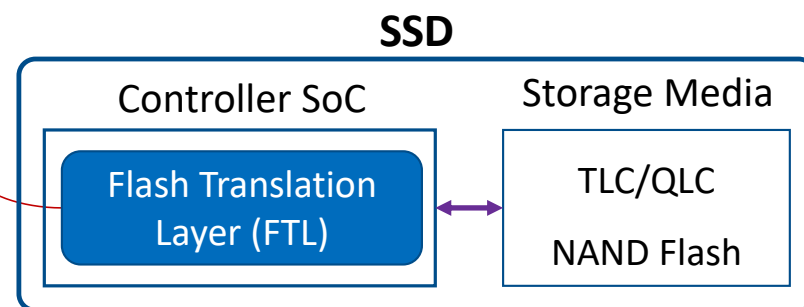


Offload the storage management task to SSDs



**Variable-length data
block management**

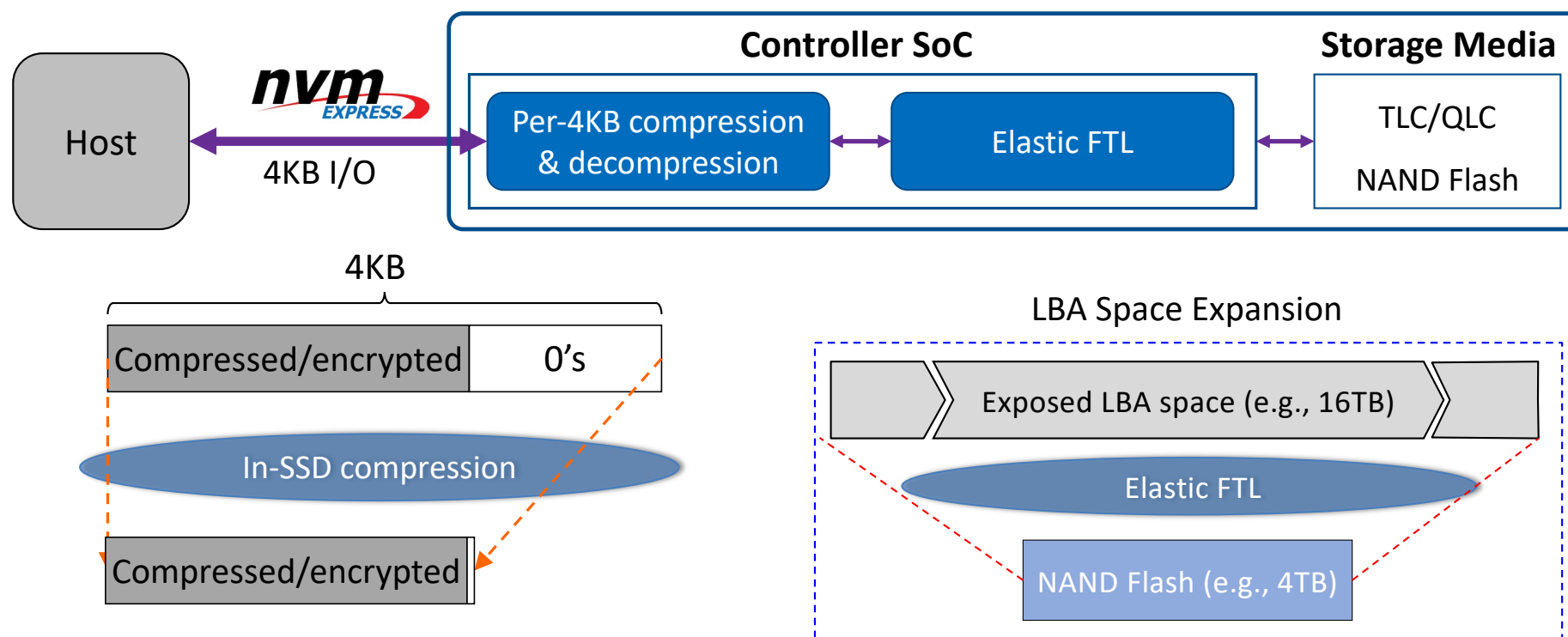
Fixed-length 4KB data
block management





SSD w/ Built-in Transparent Compression

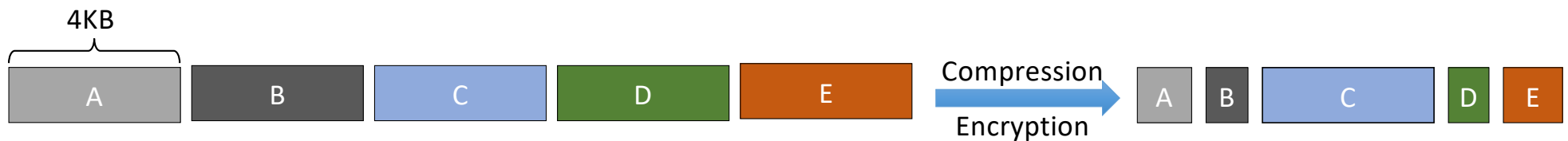
- ❑ NVMe SSD that compresses each 4KB LBA block, completely transparent to the host



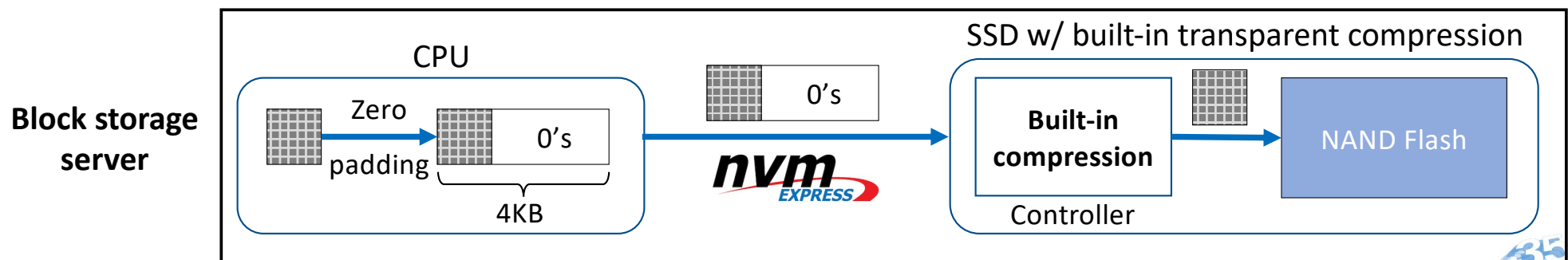
Block Storage Server Data Management



Storage management of compressed/encrypted variable-length blocks



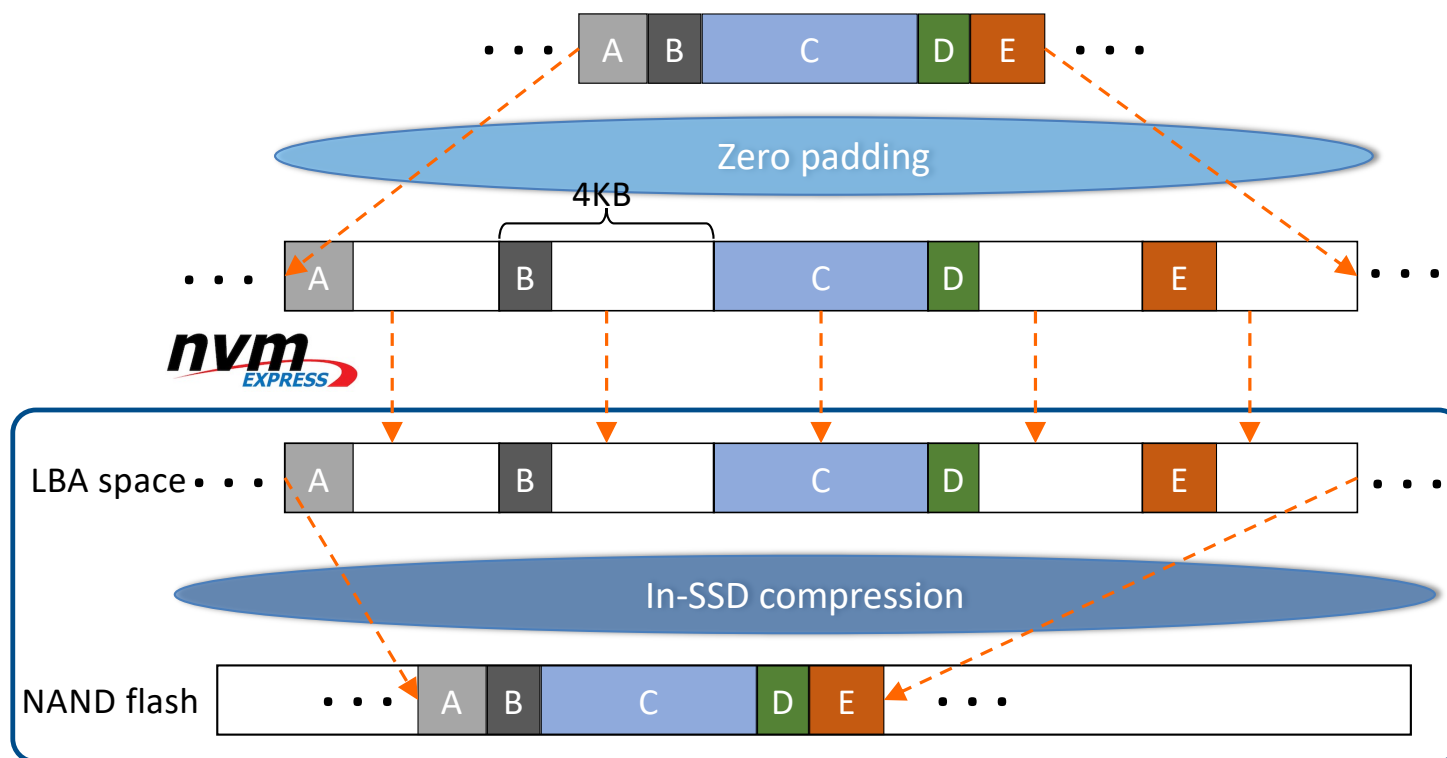
Offload the storage management task to SSDs



Block Storage Server Data Management



Flash Memory Summit



Compression ration ↑



SSD LBA space expansion ↑



FTL mapping table size ↑



Limited DRAM inside SSD

Block Storage Server Data Management

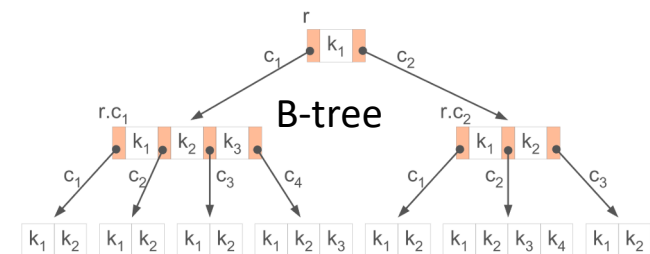
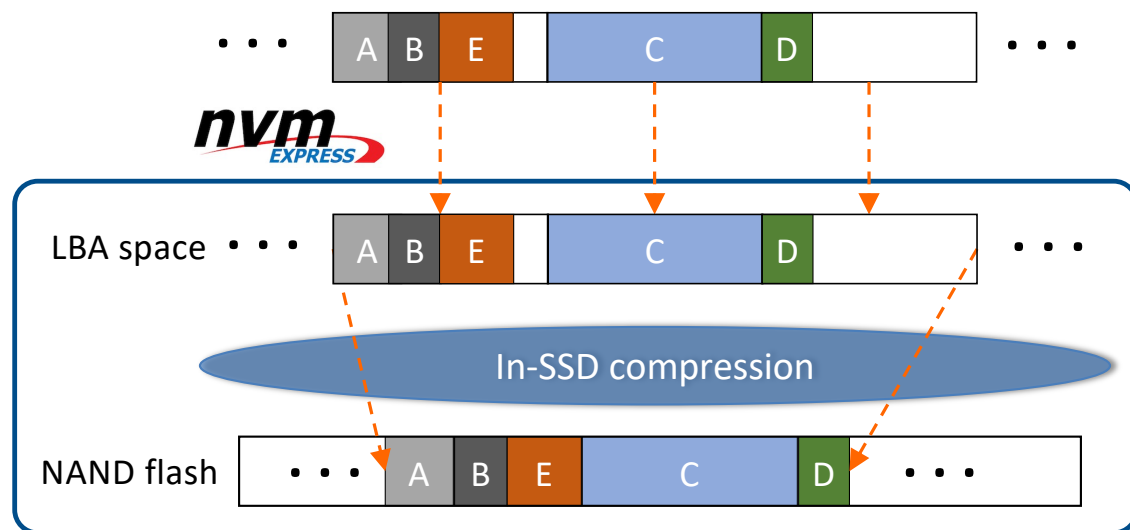


Flash Memory Summit



Adaptive data block fusion

- Categorize data blocks as write-hot and write-cold
 - Each write-hot data block entirely occupies one 4KB LBA on SSD
 - One or multiple write-cold blocks share one 4KB LBA on SSD



Manage data blocks fusion

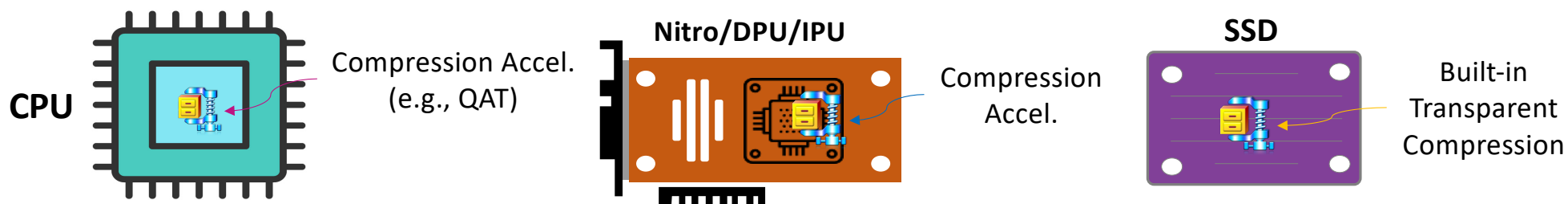


Further support snapshot on
block storage server

Conclusion



Flash Memory Summit



Disaggregated Transparent Compression over Disaggregated Infrastructure

