Private Information Retrieval

Şennur Ulukuş

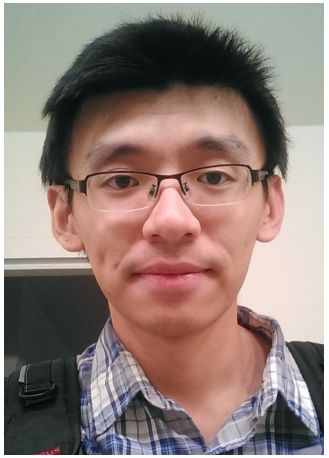
Private Information Retrieval: How to Get Something Without Revealing What You Got

Şennur Ulukuş

Private Information Retrieval: How to Get Something Without Revealing What You Got

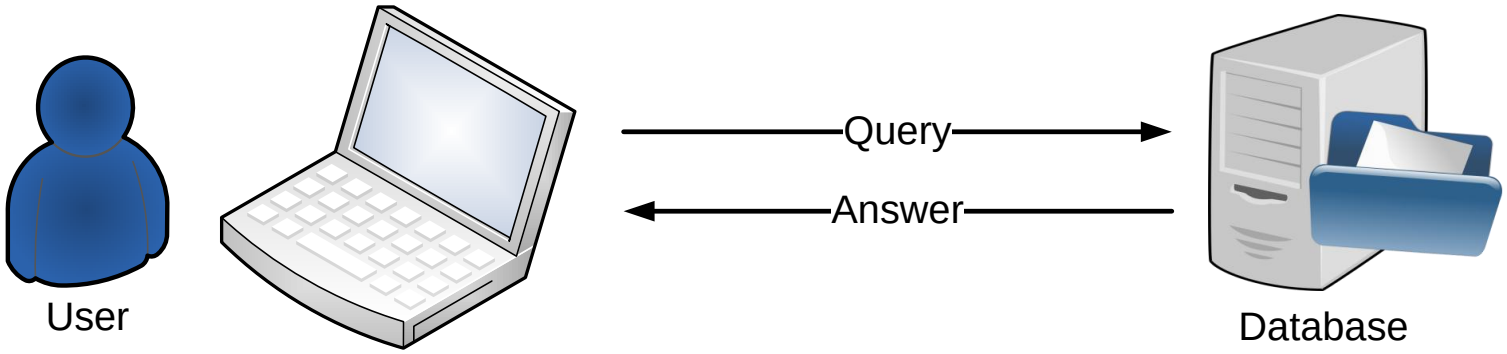
Şennur Ulukuş

Karim Banawan

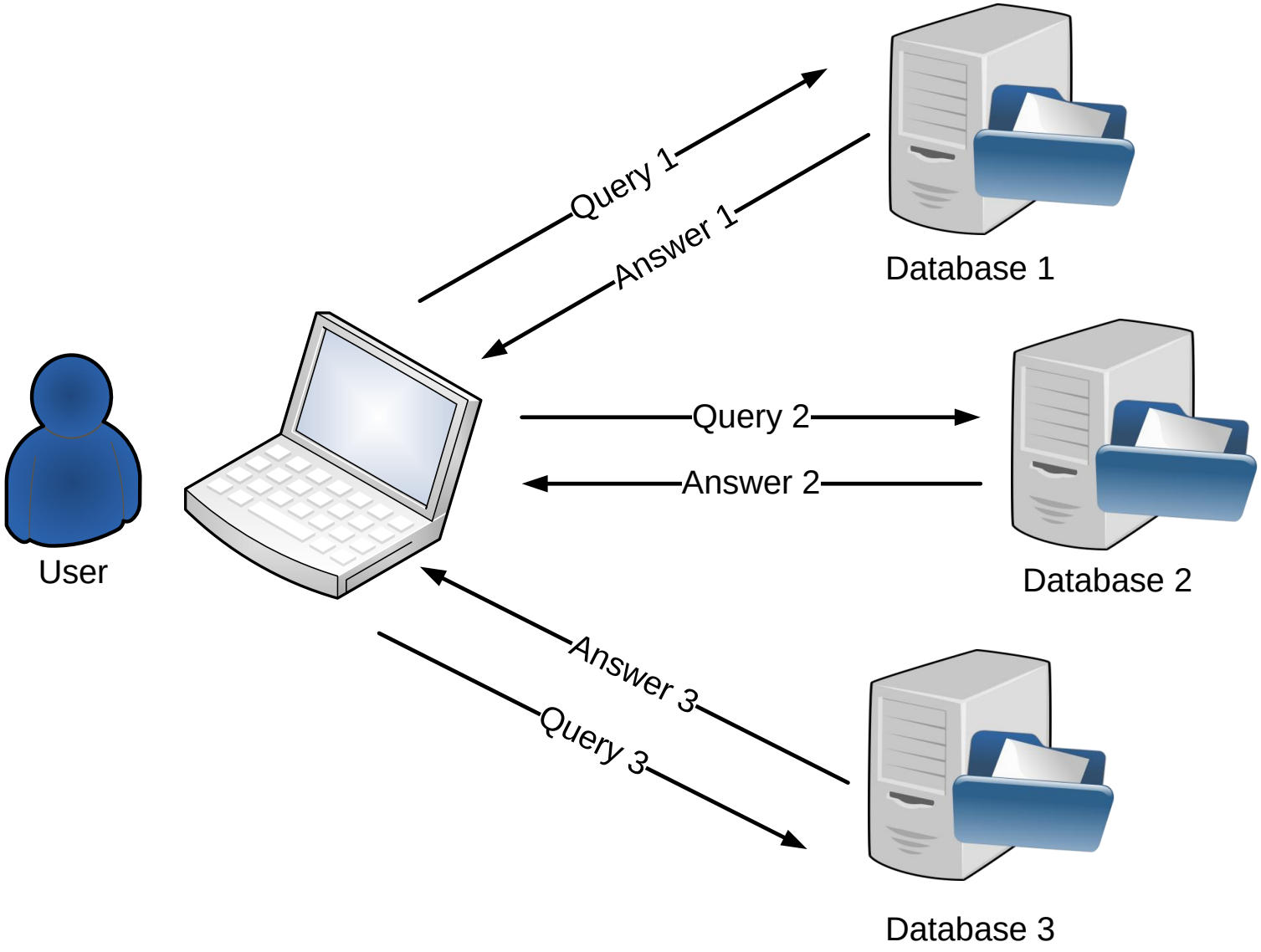


Yi-Peng Wei

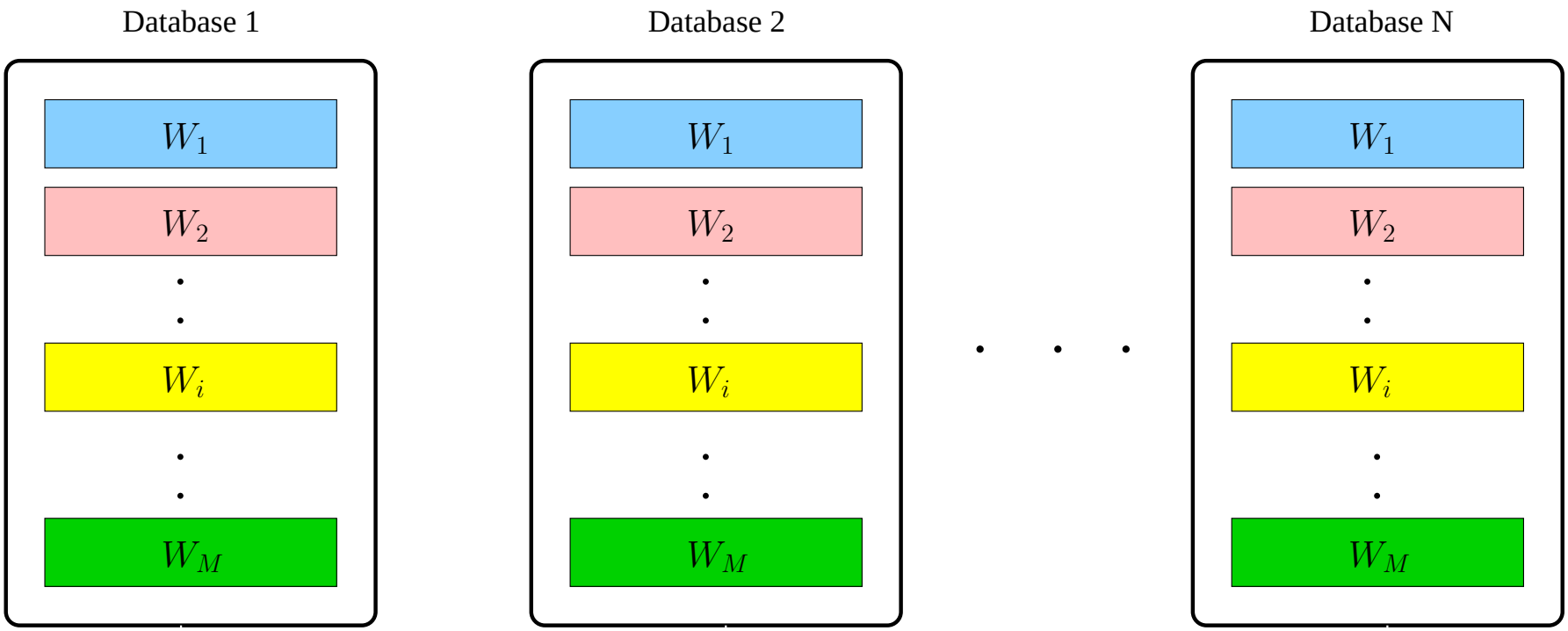
Private Information Retrieval (PIR) Problem: Single Server



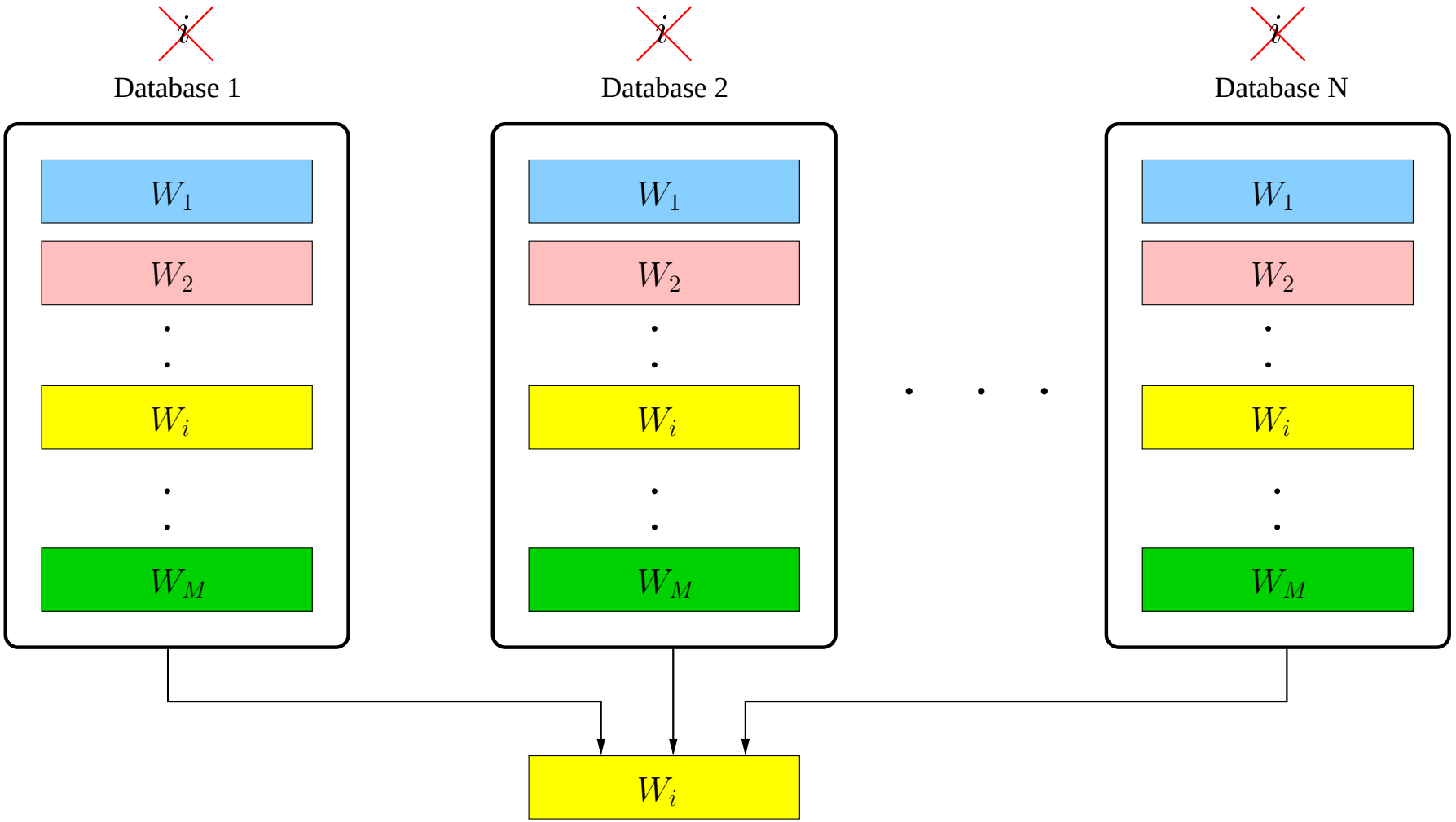
Private Information Retrieval (PIR) Problem: Multi Servers



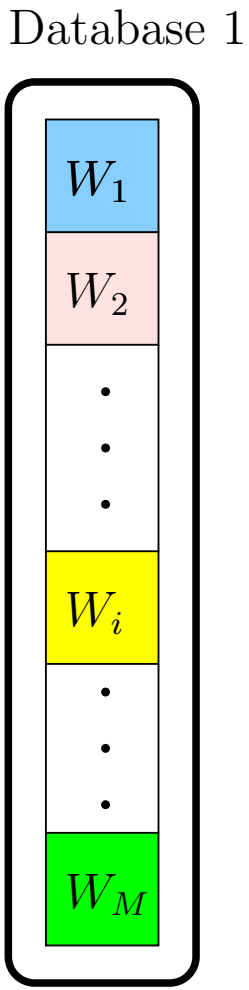
Classical PIR Model



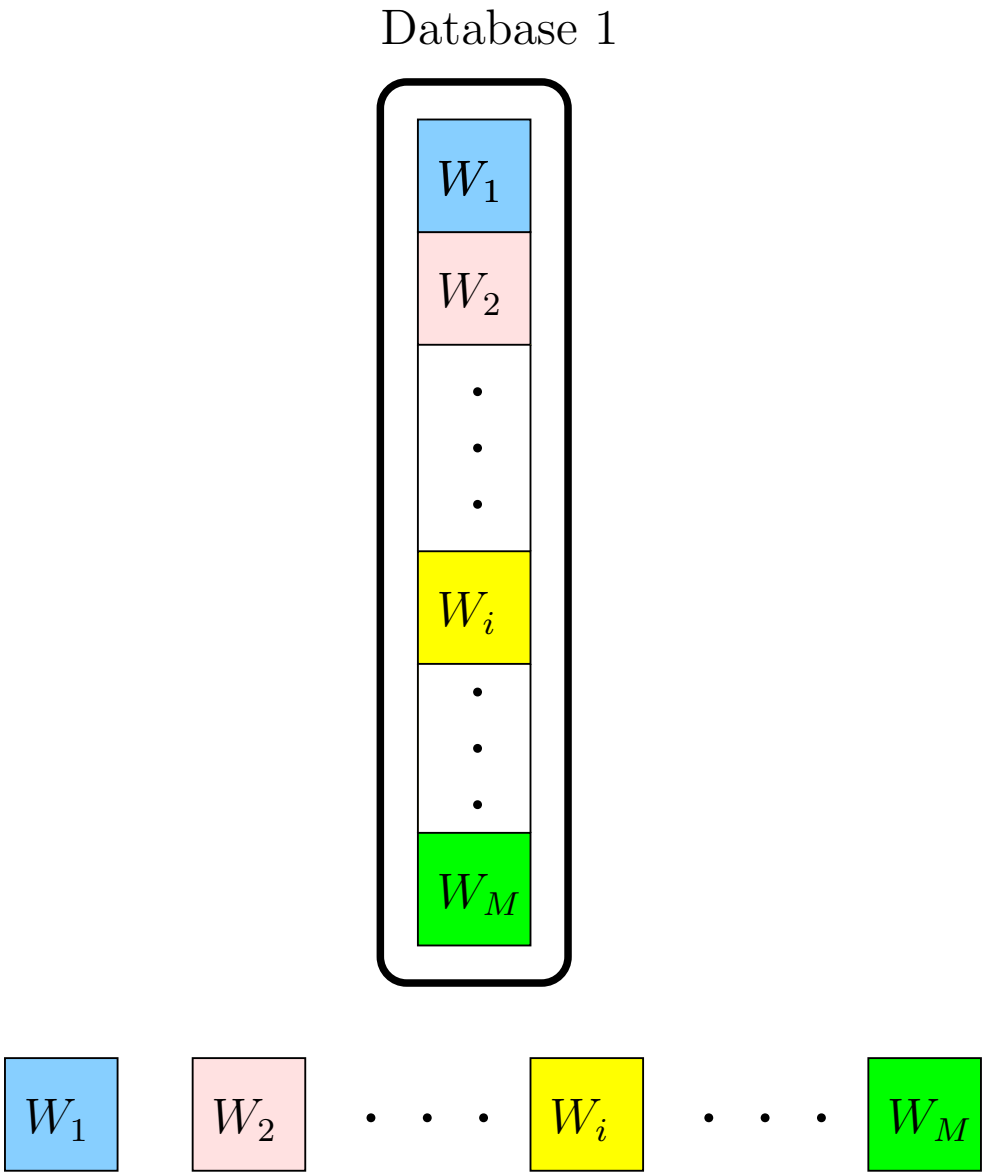
Classical PIR Model



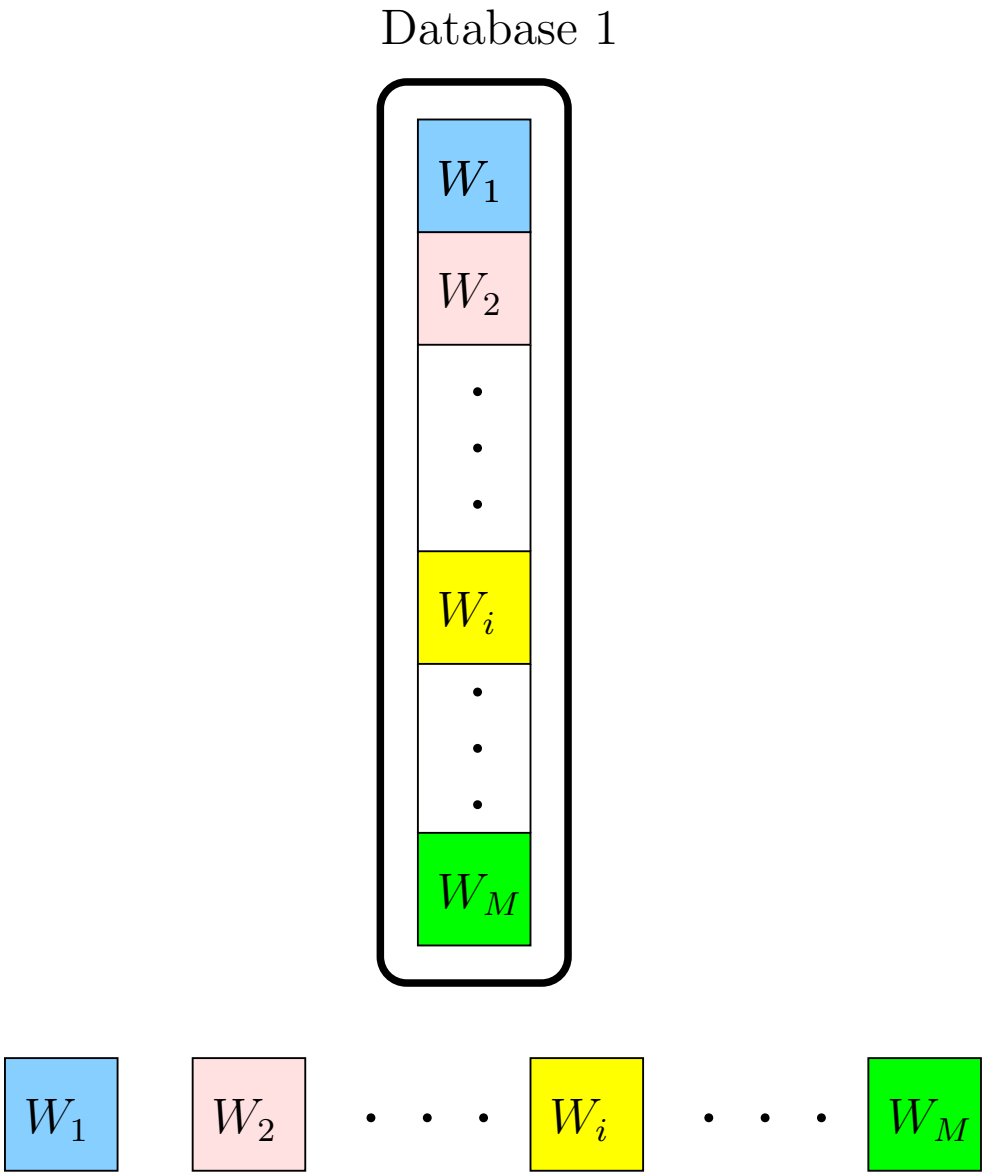
Single-Database PIR



Single-Database PIR



Single-Database PIR



- **Can we do better than $\frac{1}{M}$?**

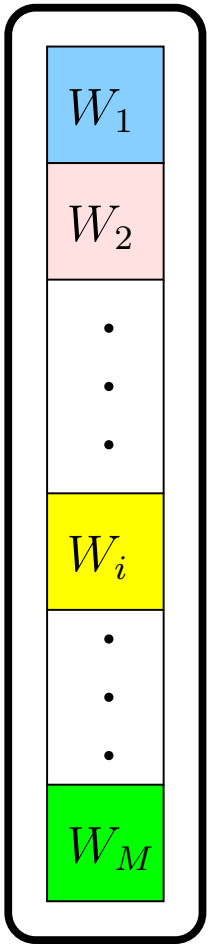
Chor et al.[1] Basic Scheme: Two-Database PIR, 1-Bit Messages

random vector $\in \{0,1\}^M$

h_1	h_2	$\cdots$	h_i	$\cdots$	h_M

$$A_1^{[i]} = \sum_{m=1}^{M} h_m W_m$$

Database 1



[1] B. Chor, E. Kushilevitz, O. Goldreich, and M. Sudan. Private information retrieval.Journal of the ACM, 45(6):965-981, 1998.

Chor et al.[1] Basic Scheme: Two-Database PIR, 1-Bit Messages

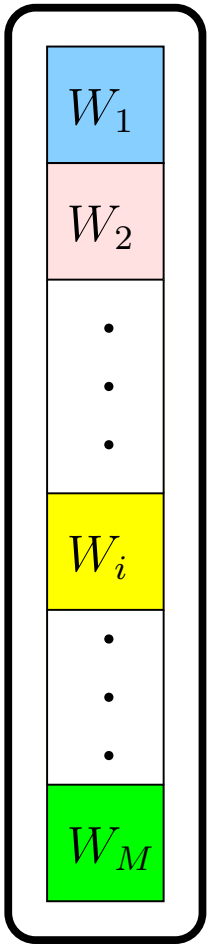
random vector $\in \{0,1\}^M$

h_1	h_2	$\cdots$	$h_i + 1$	$\cdots$	h_M

$$A_1^{[i]} = \sum_{m=1}^{M} h_m W_m$$

$$A_2^{[i]} = \sum_{m=1}^{M} h_m W_m + W_i$$

Database 2



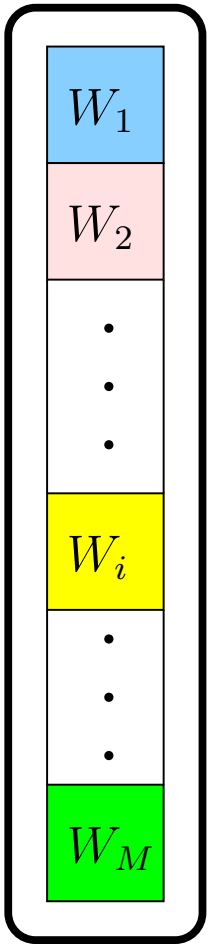
[1] B. Chor, E. Kushilevitz, O. Goldreich, and M. Sudan. Private information retrieval.Journal of the ACM, 45(6):965-981, 1998.

Chor et al.[1] Basic Scheme: Two-Database PIR, 1-Bit Messages

random vector $\in \{0,1\}^M$

h_1	h_2	$\cdots$	$h_i + 1$	$\cdots$	h_M

Database 2



$$A_1^{[i]} = \sum_{m=1}^{M} h_m W_m$$

$$A_2^{[i]} = \sum_{m=1}^{M} h_m W_m + W_i$$

$$W_i = A_2^{[i]} - A_1^{[i]}$$

[1] B. Chor, E. Kushilevitz, O. Goldreich, and M. Sudan. Private information retrieval.Journal of the ACM, 45(6):965-981, 1998.

Chor et al.[1] Basic Scheme: Two-Database PIR, 1-Bit Messages

random vector $\in \{0,1\}^M$

h_1	h_2	$\cdots$	$h_i + 1$	$\cdots$	h_M

Database 2

W_1
W_2
$\vdots$
W_i
$\vdots$
W_M

$$A_1^{[i]} = \sum_{m=1}^{M} h_m W_m$$

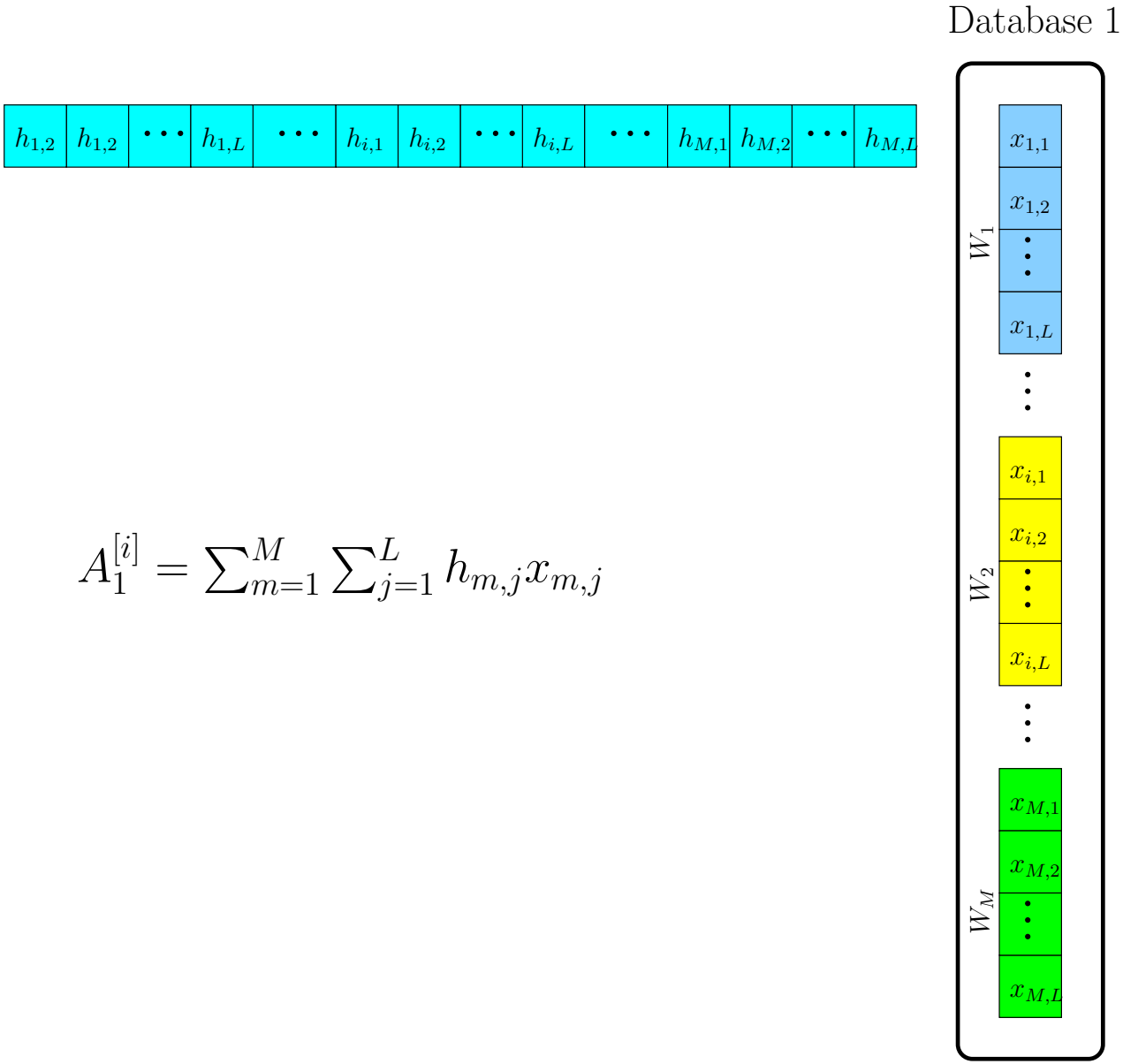
$$A_2^{[i]} = \sum_{m=1}^{M} h_m W_m + W_i$$

$$W_i = A_2^{[i]} - A_1^{[i]}$$

$$\boxed{R = \frac{1}{2}}$$

[1] B. Chor, E. Kushilevitz, O. Goldreich, and M. Sudan. Private information retrieval.Journal of the ACM, 45(6):965-981, 1998.

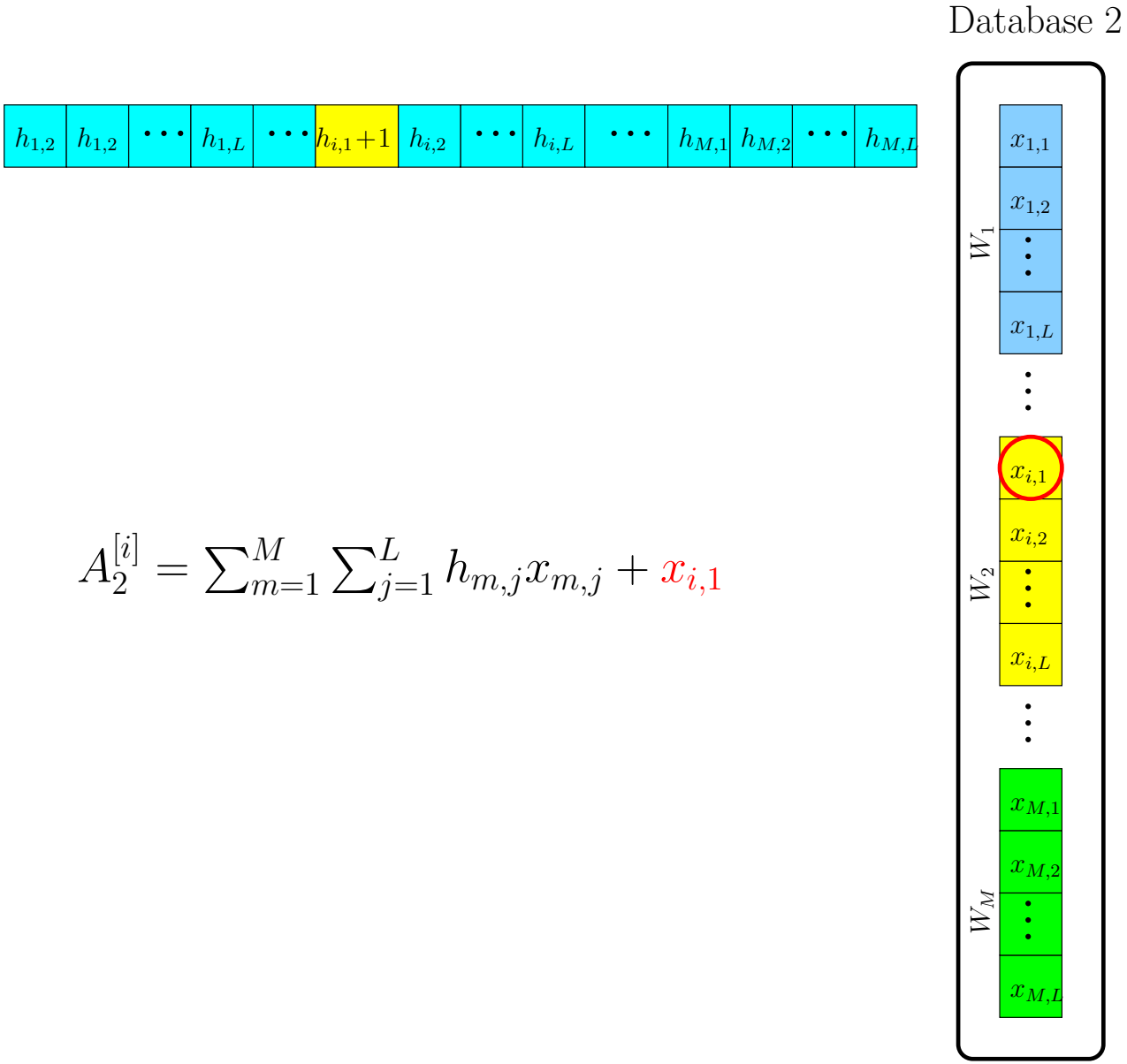
Extension for arbitrary N [Shah-Rashmi-Ramchandran][2]



$$A_1^{[i]} = \sum_{m=1}^{M} \sum_{j=1}^{L} h_{m,j} x_{m,j}$$

[2] N. B. Shah, K. V. Rashmi, and K. Ramchandran. One extra bit of download ensures perfectly private information retrieval. In IEEE ISIT, June 2014.

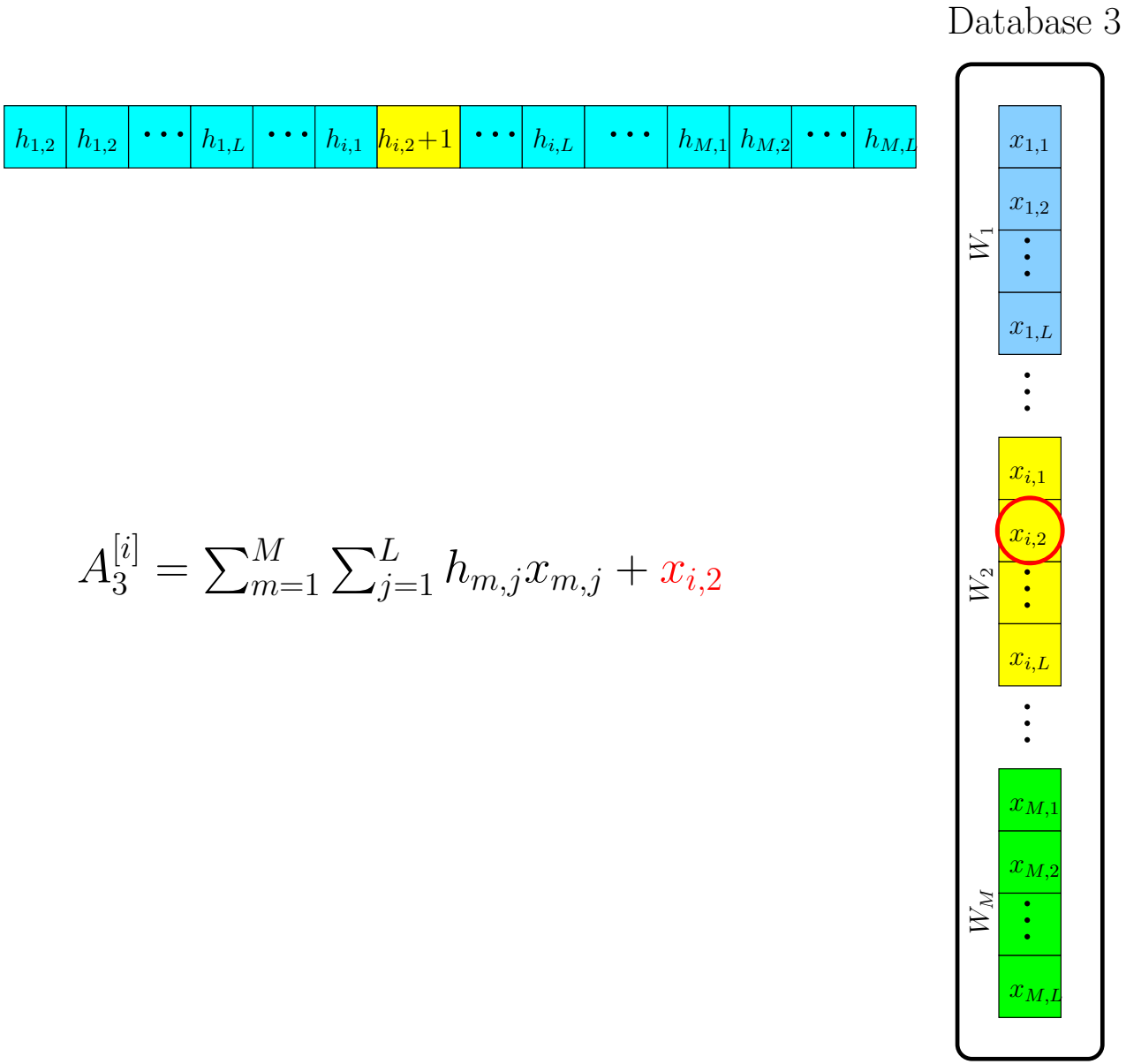
Extension for arbitrary N [Shah-Rashmi-Ramchandran][2]



$$A_2^{[i]} = \sum_{m=1}^{M} \sum_{j=1}^{L} h_{m,j} x_{m,j} + x_{i,1}$$

[2] N. B. Shah, K. V. Rashmi, and K. Ramchandran. One extra bit of download ensures perfectly private information retrieval. In IEEE ISIT, June 2014.

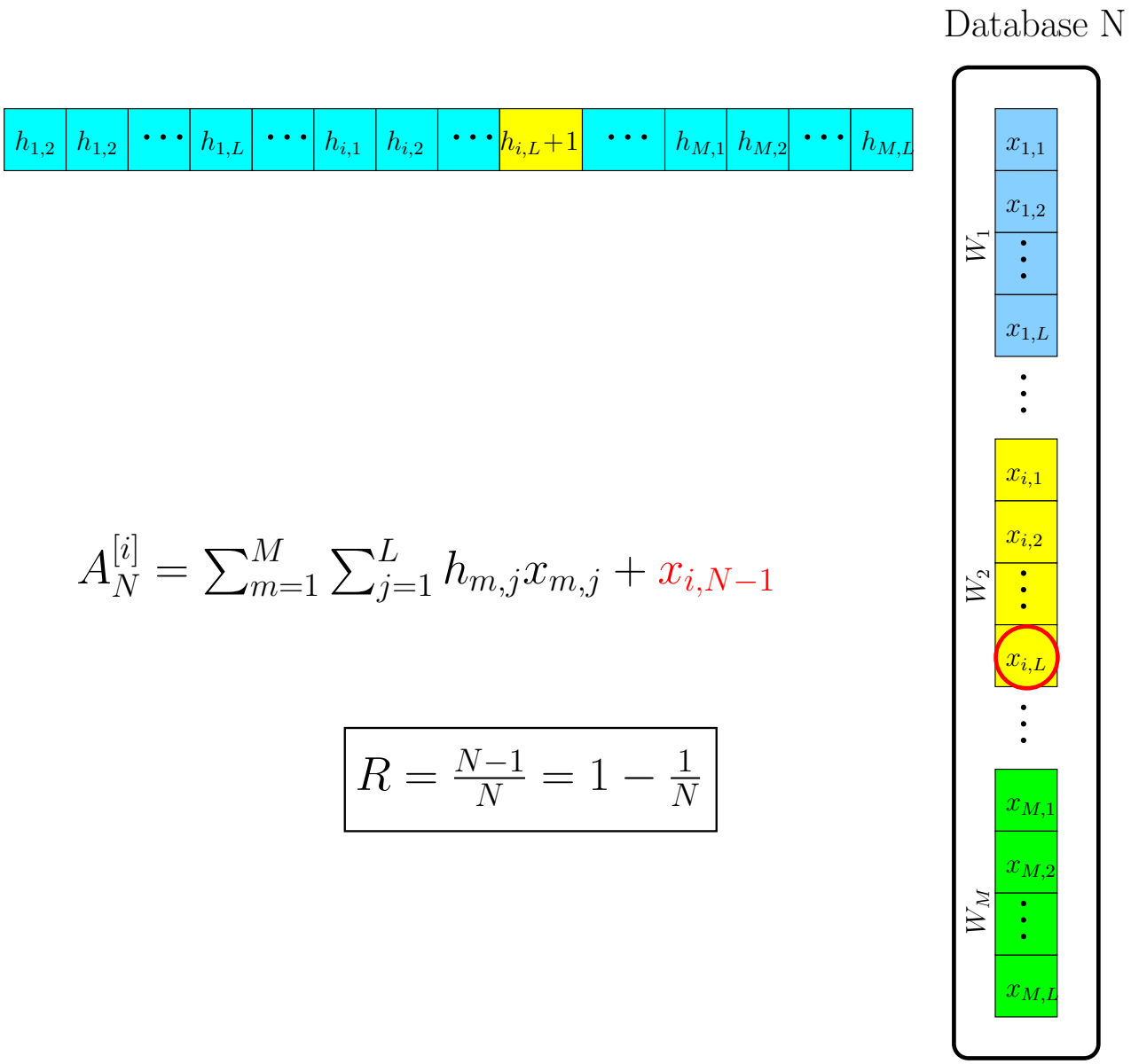
Extension for arbitrary N [Shah-Rashmi-Ramchandran][2]



$$A_3^{[i]} = \sum_{m=1}^{M} \sum_{j=1}^{L} h_{m,j} x_{m,j} + x_{i,2}$$

[2] N. B. Shah, K. V. Rashmi, and K. Ramchandran. One extra bit of download ensures perfectly private information retrieval. In IEEE ISIT, June 2014.

Extension for arbitrary N [Shah-Rashmi-Ramchandran][2]



$$A_N^{[i]} = \sum_{m=1}^{M} \sum_{j=1}^{L} h_{m,j} x_{m,j} + x_{i,N-1}$$

$$\boxed{R = \frac{N-1}{N} = 1 - \frac{1}{N}}$$

[2] N. B. Shah, K. V. Rashmi, and K. Ramchandran. One extra bit of download ensures perfectly private information retrieval. In IEEE ISIT, June 2014.

Information-Theoretic Re-Formulation of the PIR problem

- Classical PIR [Sun-Jafar].
- PIR with colluding databases (TPIR) [Sun-Jafar].
- Robust PIR (RPIR) [Sun-Jafar].
- Symmetric PIR (SPIR) [Sun-Jafar].
- Coded PIR (CPIR) [Tajeddine-El Rouayheb & Banawan-Ulukus].
- Coded symmetric PIR [Wang-Skoglund].
- PIR with arbitrary message length (LPIR) [Sun-Jafar].
- Multi-round PIR [Sun-Jafar].
- Multi-message PIR (MPIR) [Banawan-Ulukus].
- PIR with coded colluding databases [Freij-Hollanti et al. & Sun-Jafar].
- PIR with Byzantine databases (BPIR) [Banawan-Ulukus].
- Cache-aided PIR [Tandon & Wei-Banawan-Ulukus].
- PIR with PSI [Kadhe et al. & Chen-Wang-Jafar & Wei-Banawan-Ulukus].
- PIR with arbitrary collusion patterns [Tajeddine et al. & Jia-Sun-Jafar].
- Secure PIR [Wang-Skoglund].
- Private function retrieval [Sun-Jafar & Mirmohseni-Maddah-Ali & Karpuk].
- PIR under asymmetric traffic constraints [Banawan-Ulukus].
- PIR from wiretap channel II [Banawan-Ulukus].
- PIR with PSI under storage constraints [Wei-Ulukus].
- Noisy PIR (NPIR) [Banawan-Ulukus].
- PIR from multiple access channels (MAC-PIR) [Banawan-Ulukus].

Information-Theoretic Re-Formulation of the PIR problem

- Classical PIR [Sun-Jafar].
- PIR with colluding databases (TPIR) [Sun-Jafar].
- Robust PIR (RPIR) [Sun-Jafar].
- Symmetric PIR (SPIR) [Sun-Jafar].
- Coded PIR (CPIR) [Tajeddine-El Rouayheb & Banawan-Ulukus].
- Coded symmetric PIR [Wang-Skoglund].
- PIR with arbitrary message length (LPIR) [Sun-Jafar].
- Multi-round PIR [Sun-Jafar].
- Multi-message PIR (MPIR) [Banawan-Ulukus].
- PIR with coded colluding databases [Freij-Hollanti et al. & Sun-Jafar].
- PIR with Byzantine databases (BPIR) [Banawan-Ulukus].
- Cache-aided PIR [Tandon & Wei-Banawan-Ulukus].
- PIR with PSI [Kadhe et al. & Chen-Wang-Jafar & Wei-Banawan-Ulukus].
- PIR with arbitrary collusion patterns [Tajeddine et al. & Jia-Sun-Jafar].
- Secure PIR [Wang-Skoglund].
- Private function retrieval [Sun-Jafar & Mirmohseni-Maddah-Ali & Karpuk].
- PIR under asymmetric traffic constraints [Banawan-Ulukus].
- PIR from wiretap channel II [Banawan-Ulukus].
- PIR with PSI under storage constraints [Wei-Ulukus].
- Noisy PIR (NPIR) [Banawan-Ulukus].
- PIR from multiple access channels (MAC-PIR) [Banawan-Ulukus].

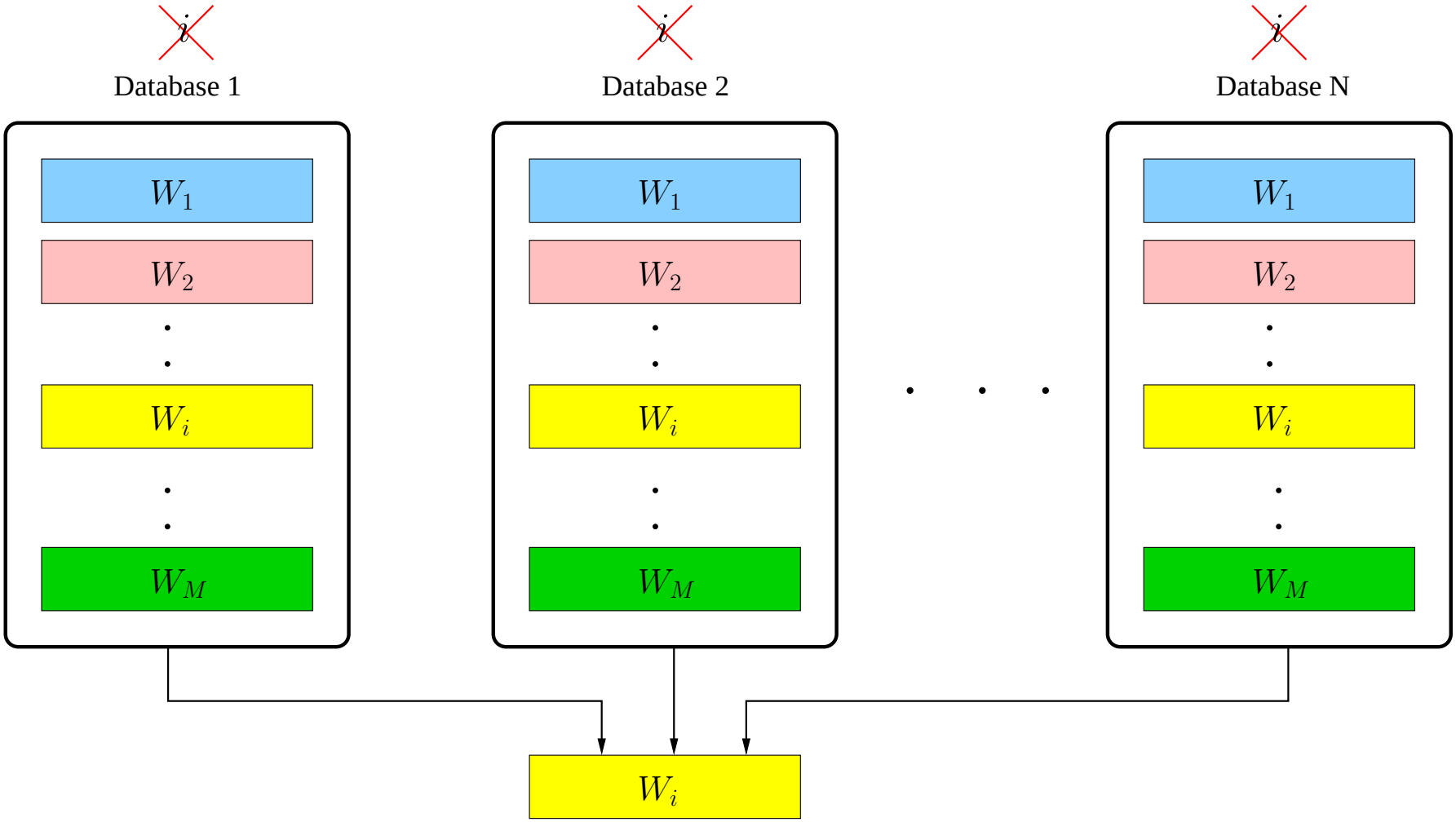
Information-Theoretic Re-Formulation of the PIR problem

- Classical PIR [Sun-Jafar].
- PIR with colluding databases (TPIR) [Sun-Jafar].
- Robust PIR (RPIR) [Sun-Jafar].
- Symmetric PIR (SPIR) [Sun-Jafar].
- Coded PIR (CPIR) [Tajeddine-El Rouayheb & Banawan-Ulukus].
- Coded symmetric PIR [Wang-Skoglund].
- PIR with arbitrary message length (LPIR) [Sun-Jafar].
- Multi-round PIR [Sun-Jafar].
- Multi-message PIR (MPIR) [Banawan-Ulukus].
- PIR with coded colluding databases [Freij-Hollanti et al. & Sun-Jafar].
- PIR with Byzantine databases (BPIR) [Banawan-Ulukus].
- Cache-aided PIR [Tandon & Wei-Banawan-Ulukus].
- PIR with PSI [Kadhe et al. & Chen-Wang-Jafar & Wei-Banawan-Ulukus].
- PIR with arbitrary collusion patterns [Tajeddine et al. & Jia-Sun-Jafar].
- Secure PIR [Wang-Skoglund].
- Private function retrieval [Sun-Jafar & Mirmohseni-Maddah-Ali & Karpuk].
- PIR under asymmetric traffic constraints [Banawan-Ulukus].
- PIR from wiretap channel II [Banawan-Ulukus].
- PIR with PSI under storage constraints [Wei-Ulukus].
- Noisy PIR (NPIR) [Banawan-Ulukus].
- PIR from multiple access channels (MAC-PIR) [Banawan-Ulukus].

Information-Theoretic Re-Formulation of the PIR problem

- Classical PIR [Sun-Jafar].
- PIR with colluding databases (TPIR) [Sun-Jafar].
- Robust PIR (RPIR) [Sun-Jafar].
- Symmetric PIR (SPIR) [Sun-Jafar].
- Coded PIR (CPIR) [Tajeddine-El Rouayheb & Banawan-Ulukus].
- Coded symmetric PIR [Wang-Skoglund].
- PIR with arbitrary message length (LPIR) [Sun-Jafar].
- Multi-round PIR [Sun-Jafar].
- Multi-message PIR (MPIR) [Banawan-Ulukus].
- PIR with coded colluding databases [Freij-Hollanti et al. & Sun-Jafar].
- PIR with Byzantine databases (BPIR) [Banawan-Ulukus].
- Cache-aided PIR [Tandon & Wei-Banawan-Ulukus]
- PIR with PSI [Kadhe et al. & Chen-Wang-Jafar & Wei-Banawan-Ulukus].
- PIR with arbitrary collusion patterns [Tajeddine et al. & Jia-Sun-Jafar].
- Secure PIR [Wang-Skoglund].
- Private function retrieval [Sun-Jafar & Mirmohseni-Maddah-Ali & Karpuk].
- PIR under asymmetric traffic constraints [Banawan-Ulukus].
- PIR from wiretap channel II [Banawan-Ulukus].
- PIR with PSI under storage constraints [Wei-Ulukus].
- Noisy PIR (NPIR) [Banawan-Ulukus].
- PIR from multiple access channels (MAC-PIR) [Banawan-Ulukus].

Classical PIR Model [Sun-Jafar][3]



[3] H. Sun and S. A. Jafar, The Capacity of Private Information Retrieval, in IEEE Transactions on Information Theory, vol. 63, no. 7, pp. 4075-4088, July 2017.

Formal Formulation

- Queries and messages are independent

$$I(Q_1^{[i]}, \cdots, Q_N^{[i]}; W_1, \cdots, W_M) = 0$$

- Answers are fully determined by messages and queries

$$H(A_n^{[i]} | Q_n^{[i]}, W_1, \cdots, W_M) = 0, \quad n \in \{1, \cdots, N\}$$

- Reliability constraint

$$H(W_i | A_1^{[i]}, \cdots, A_N^{[i]}, Q_1^{[i]}, \cdots, Q_N^{[i]}) = o(L)$$

- Privacy constraint

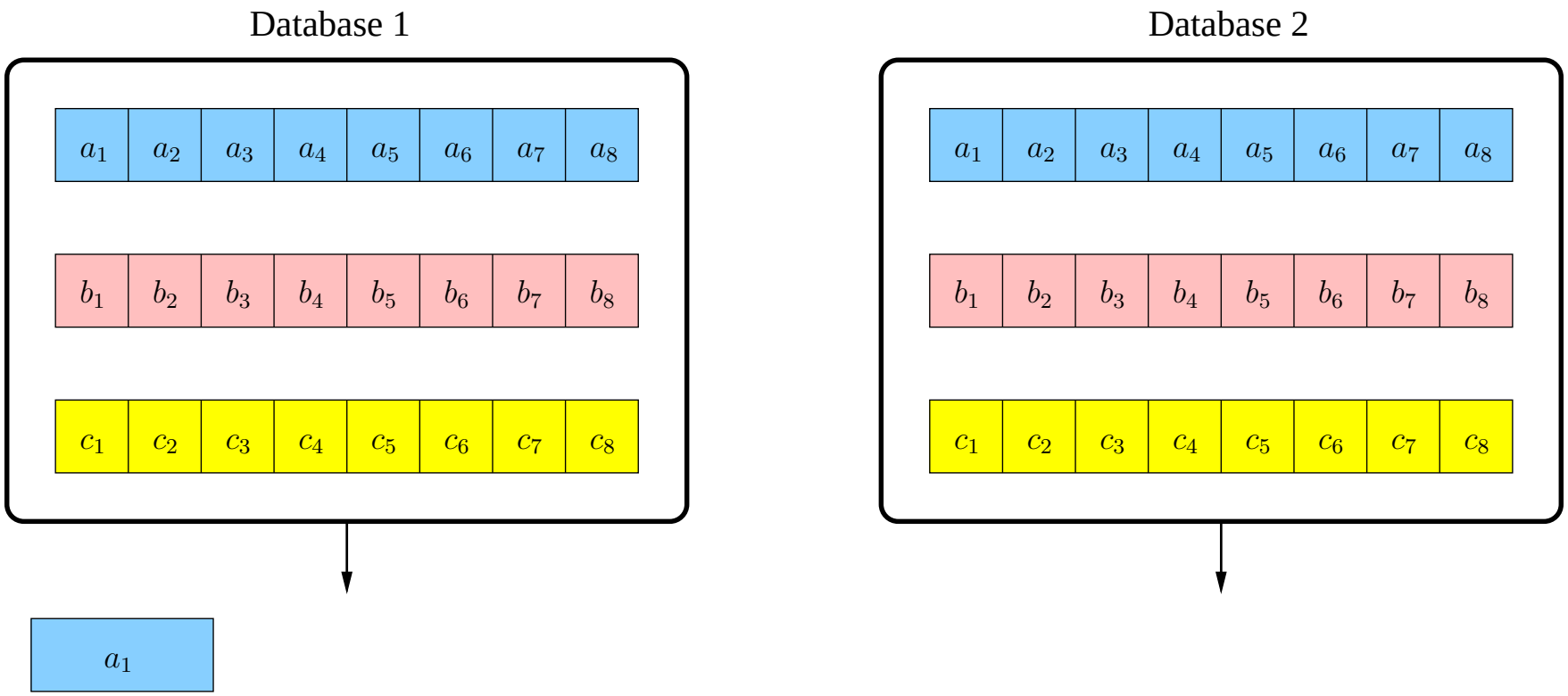
$$I(Q_n^{[i]}; i) = 0, \quad n \in \{1, \cdots, N\}$$

- Retrieval rate

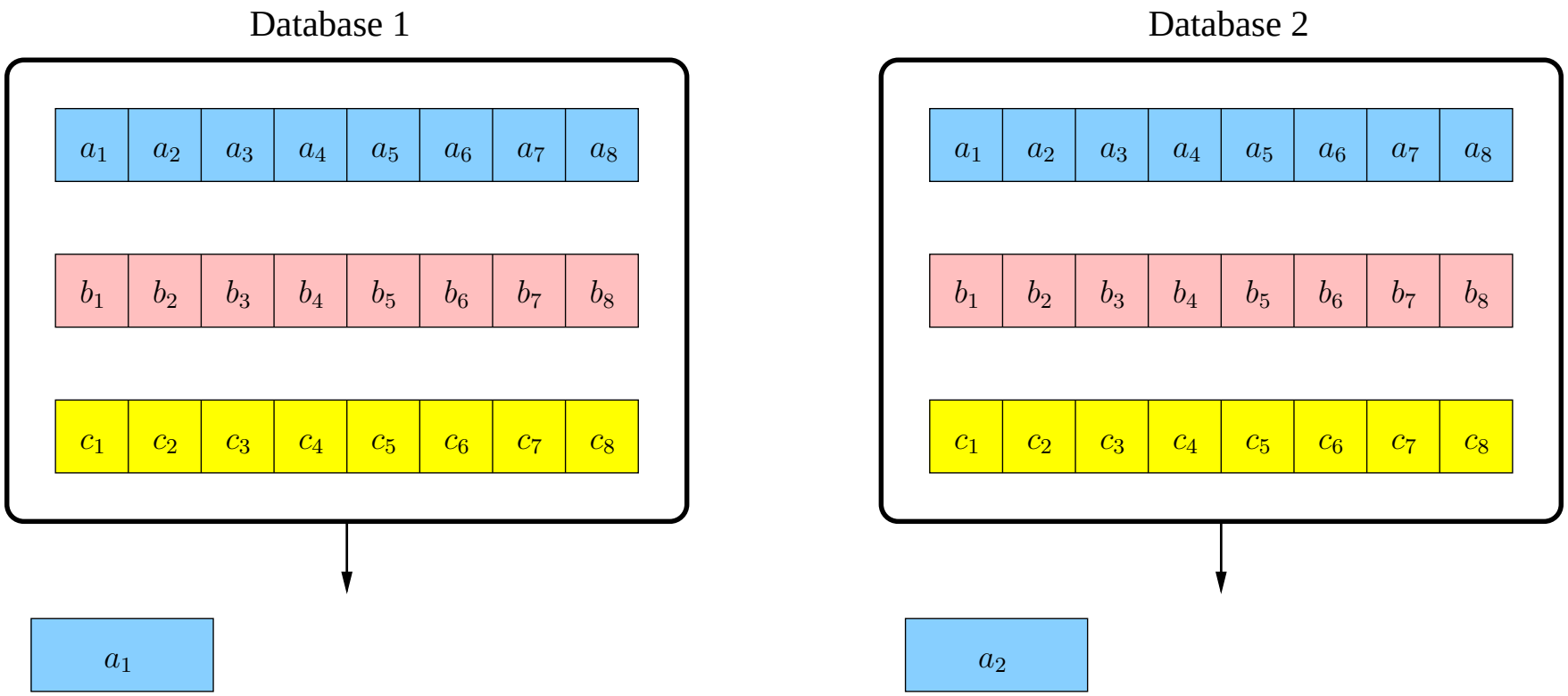
$$R = \frac{H(W_i)}{\sum_{n=1}^{N} H(A_n^{[i]})}$$

- The PIR capacity C is the supremum of R over all retrieval schemes.

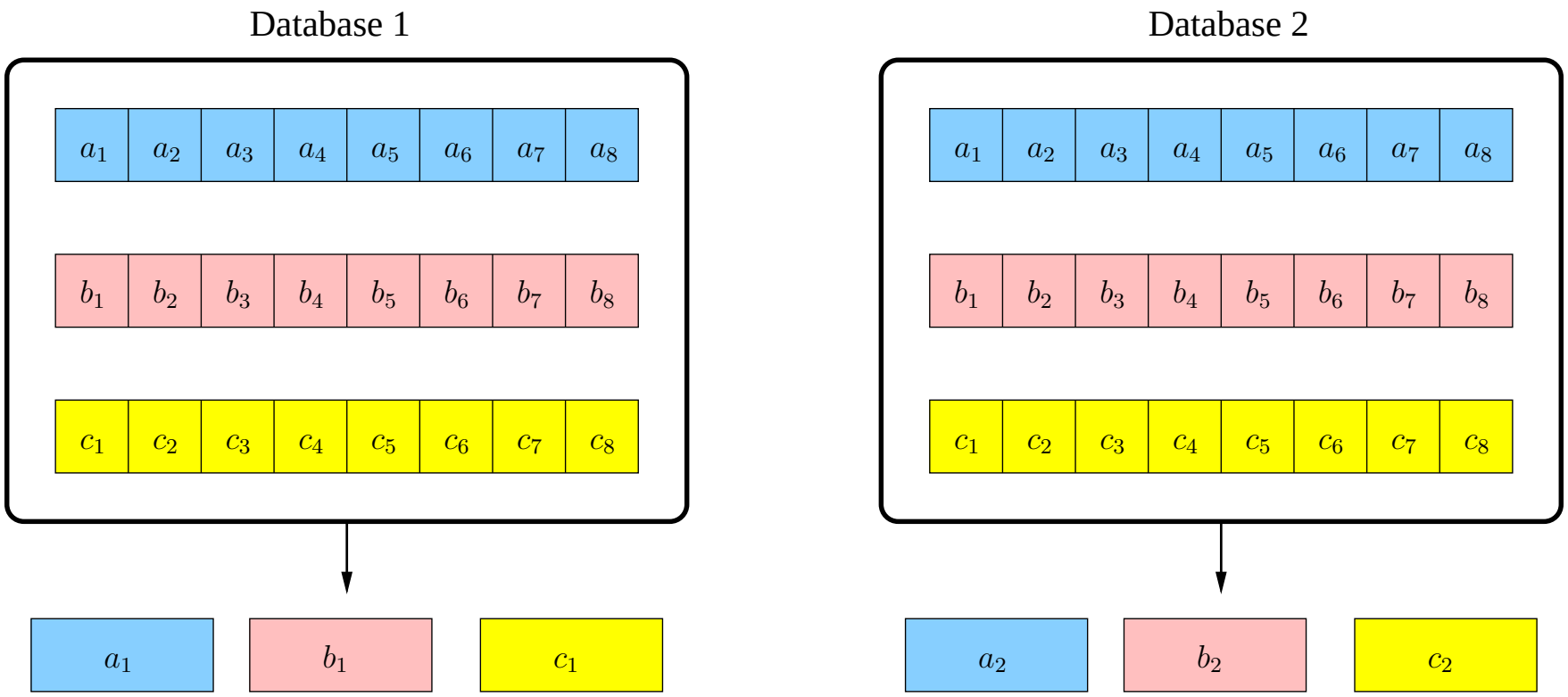
Example $M = 3, N = 2$: Retrieval



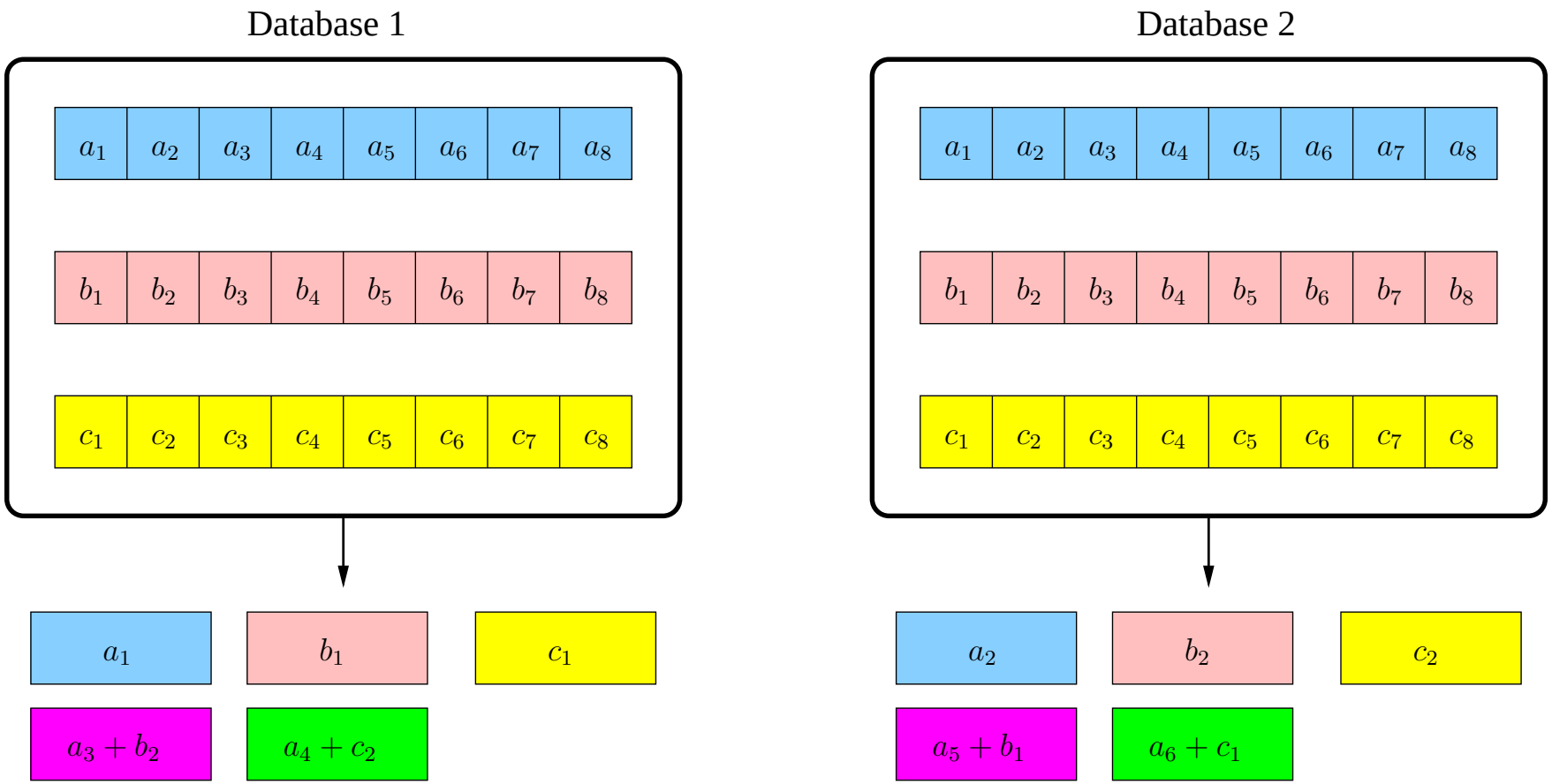
Example $M = 3, N = 2$: Retrieval



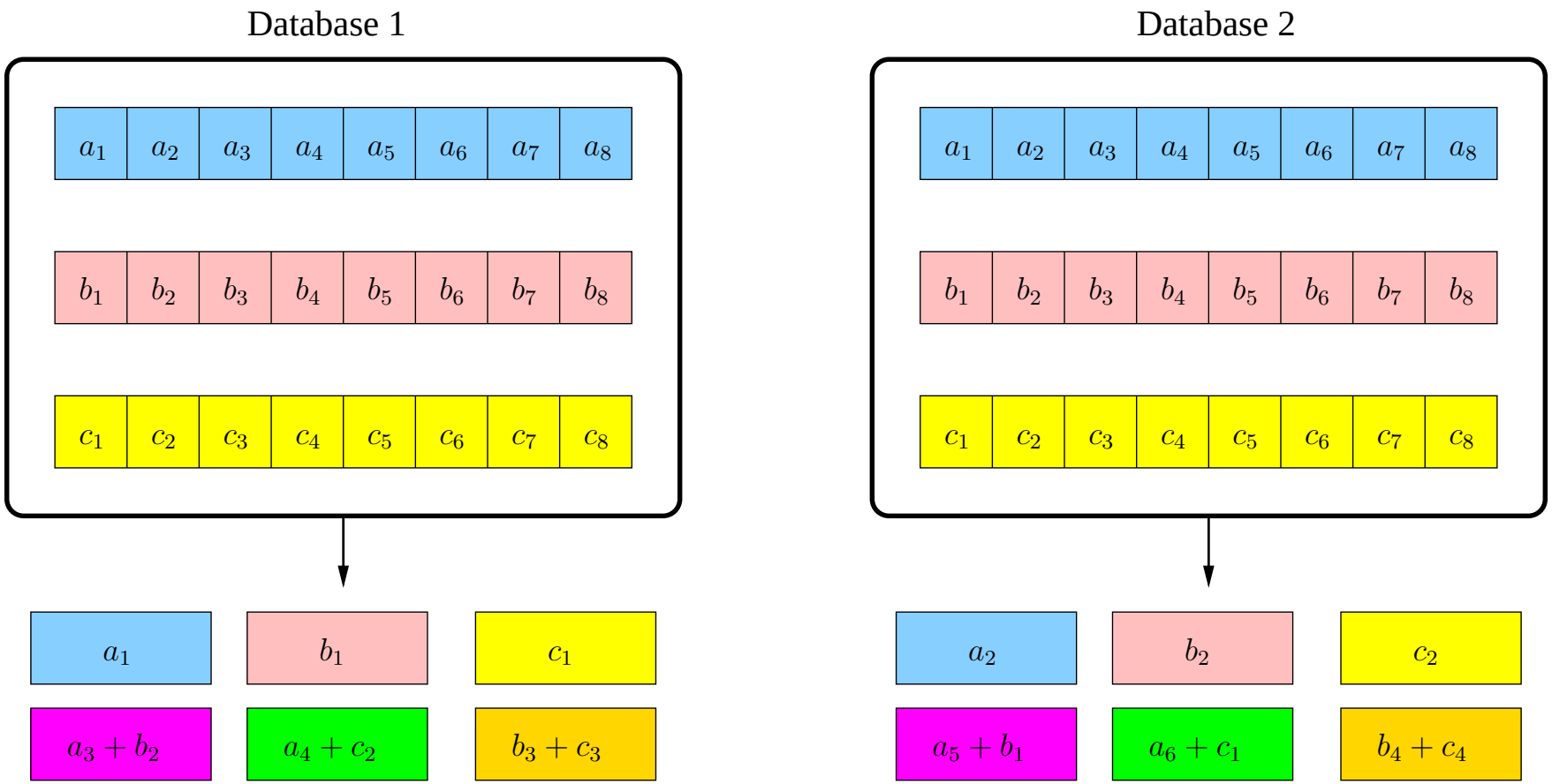
Example $M = 3, N = 2$: Retrieval



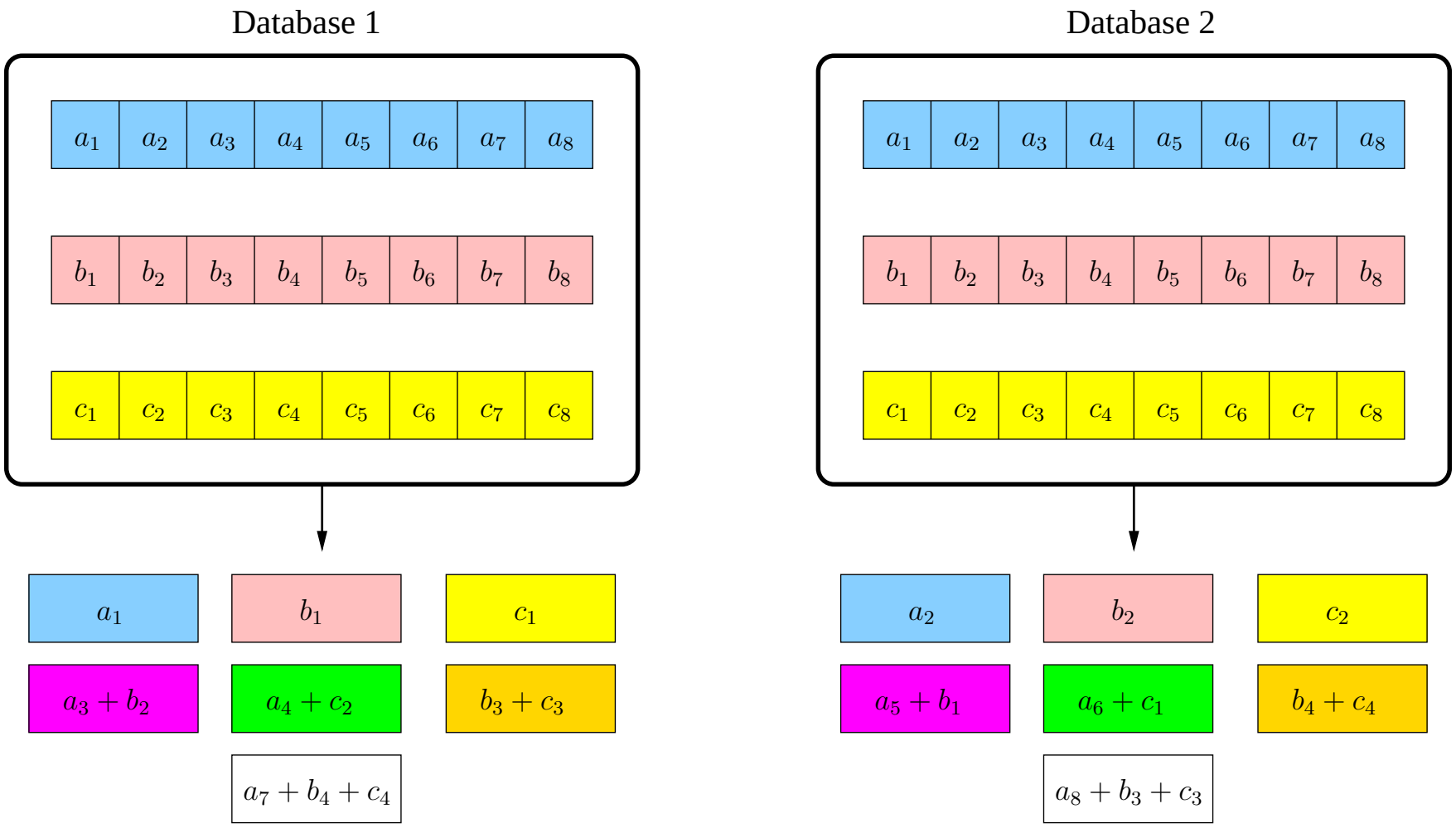
Example $M = 3, N = 2$: Retrieval



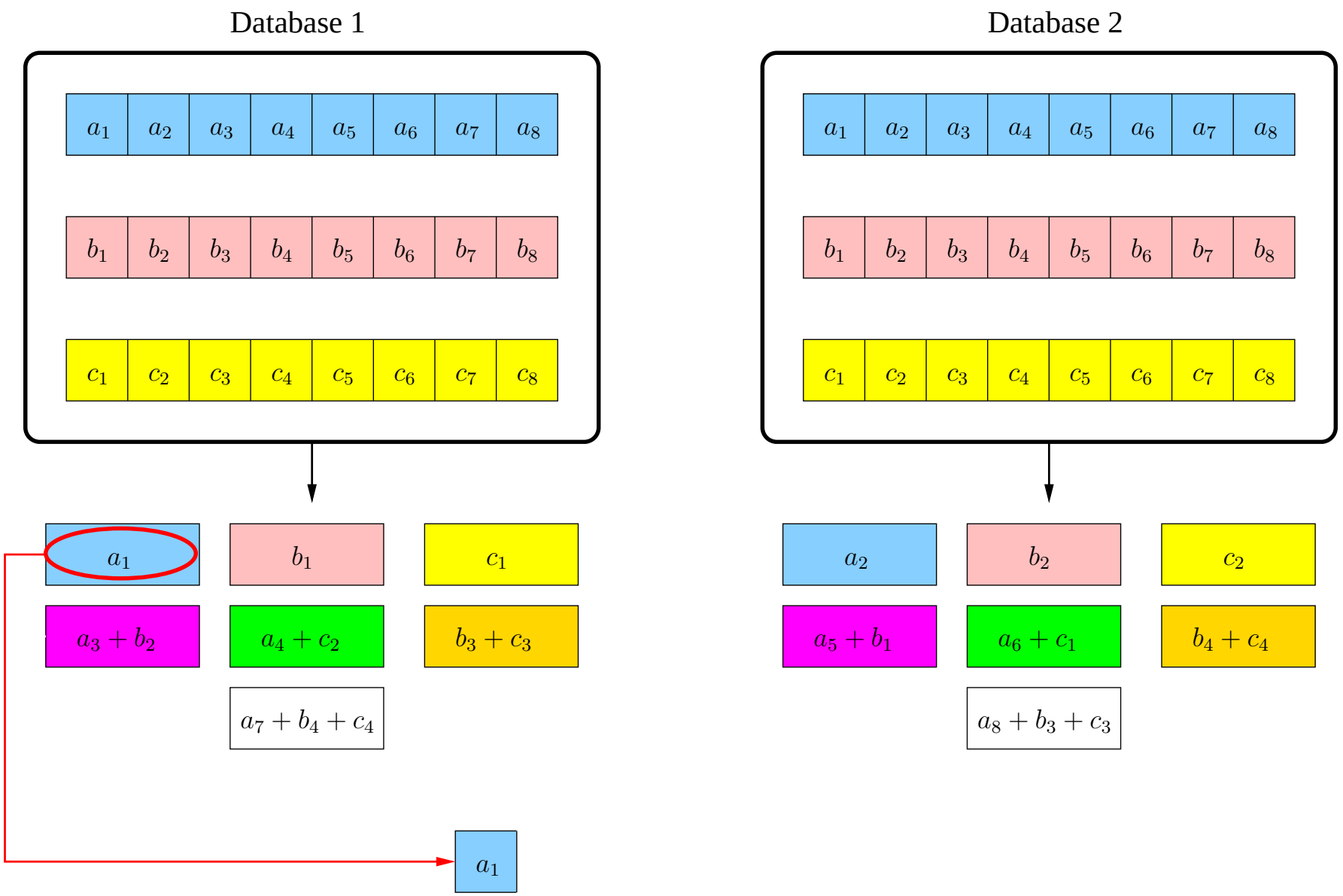
Example $M = 3, N = 2$: Retrieval



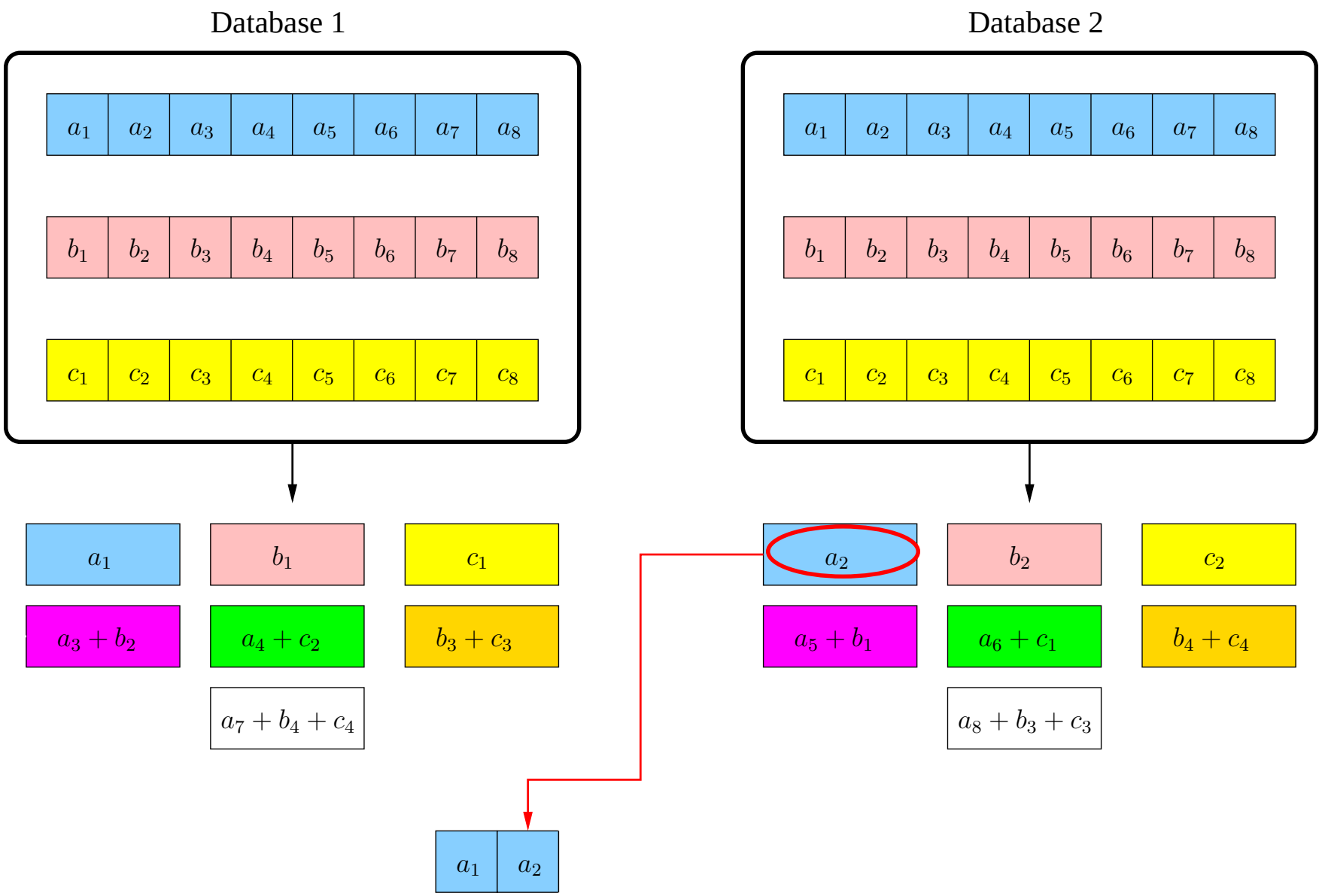
Example $M = 3, N = 2$: Retrieval



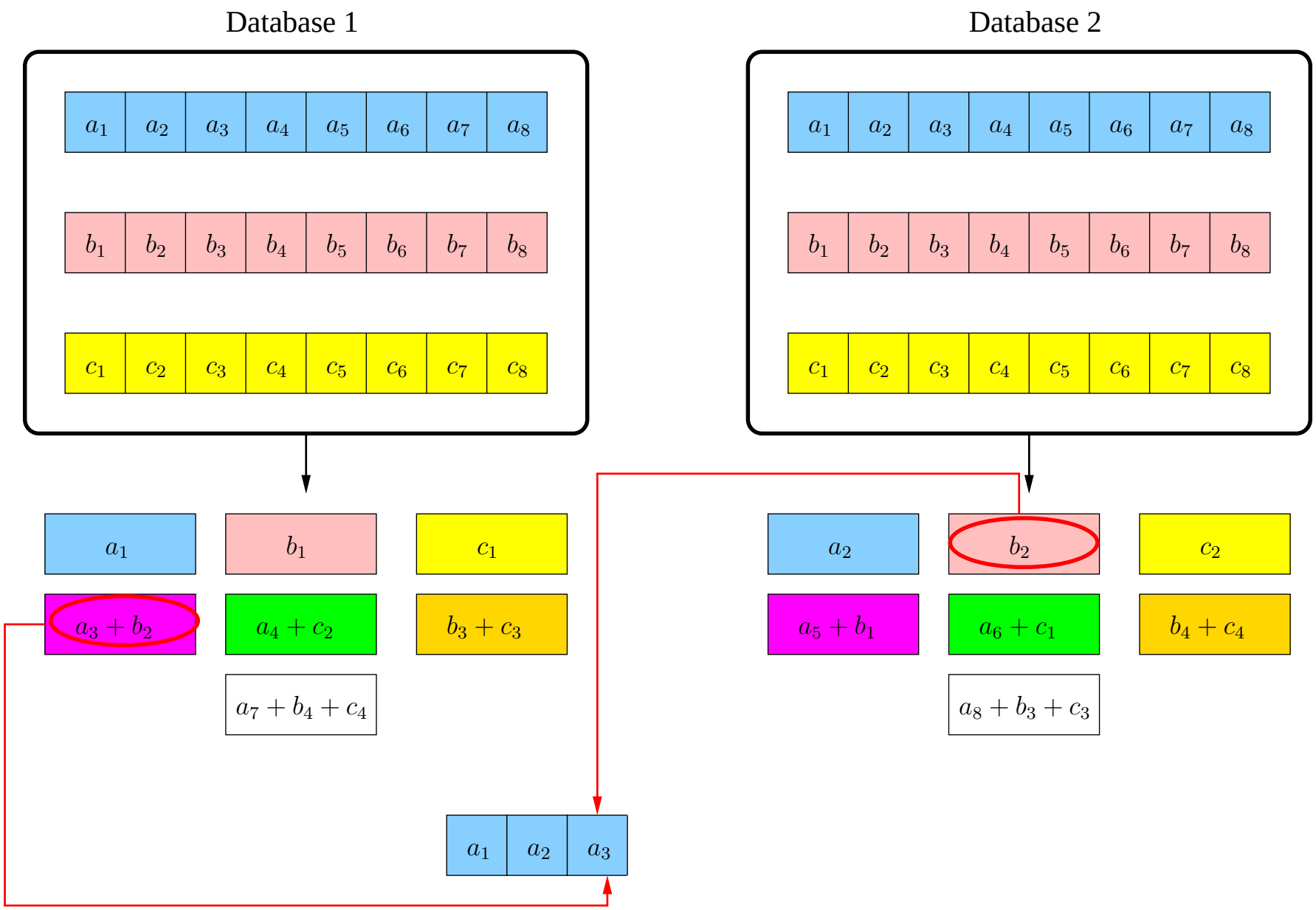
Example $M = 3, N = 2$: Decoding



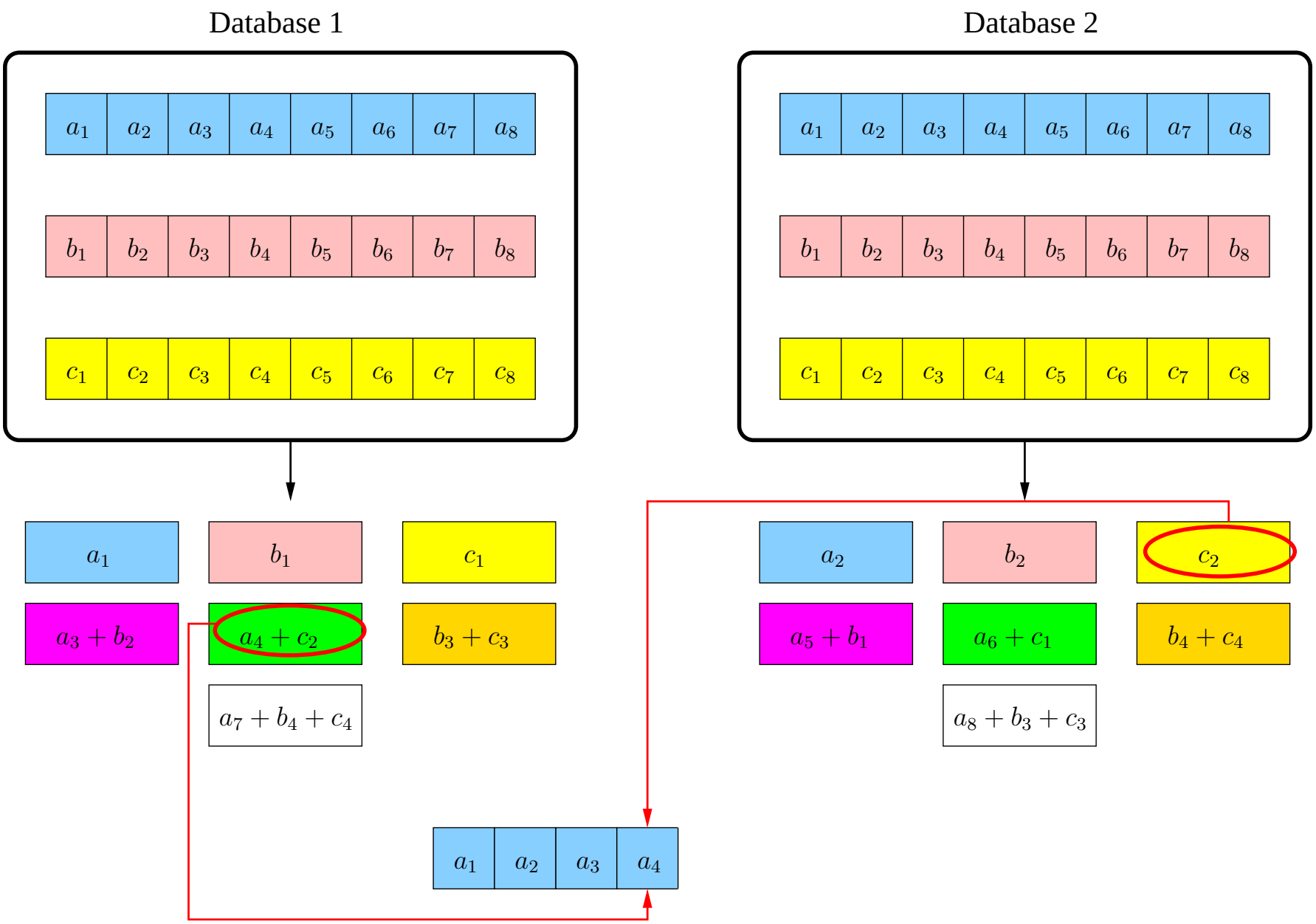
Example $M = 3, N = 2$: Decoding



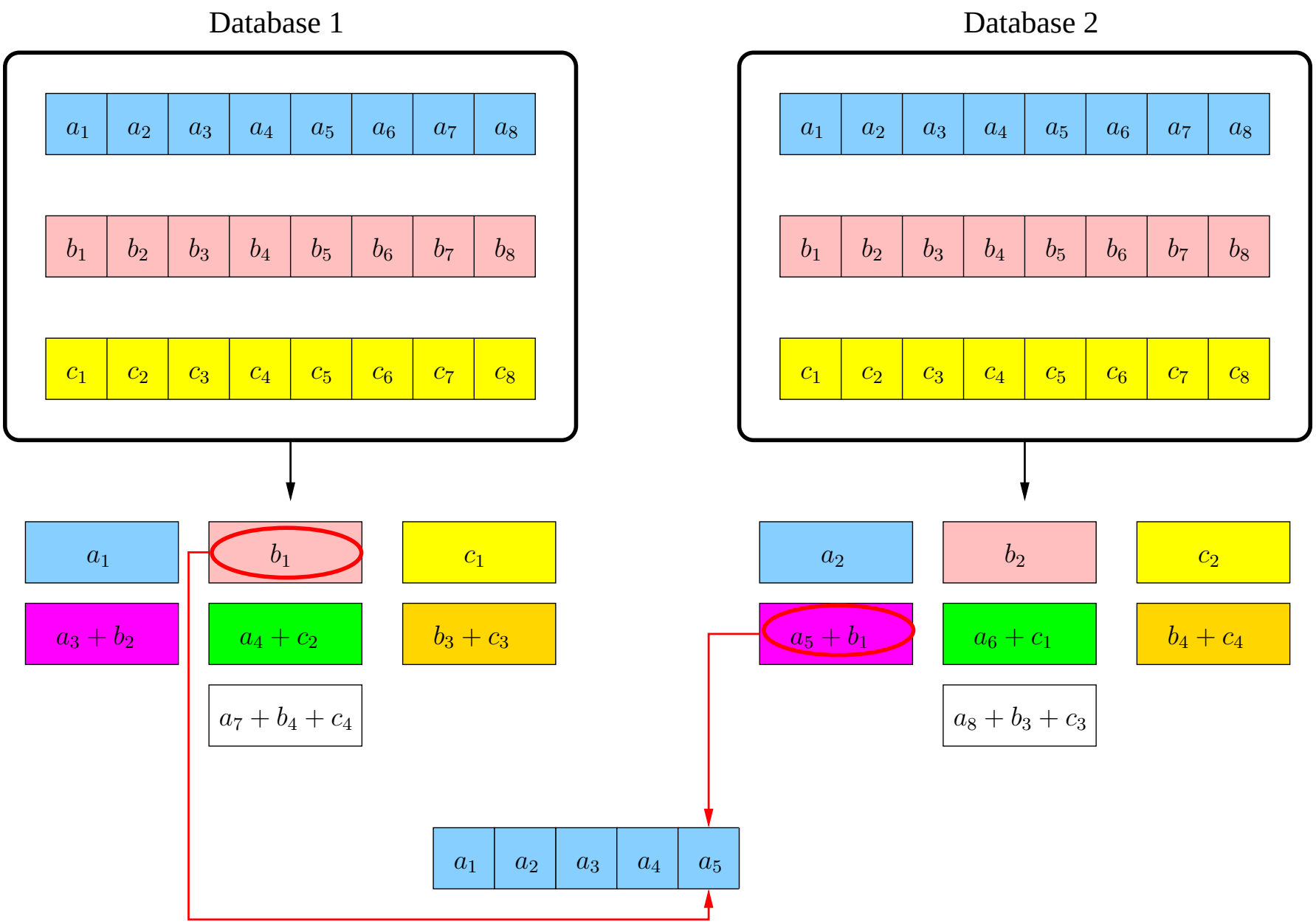
Example $M = 3, N = 2$: Decoding



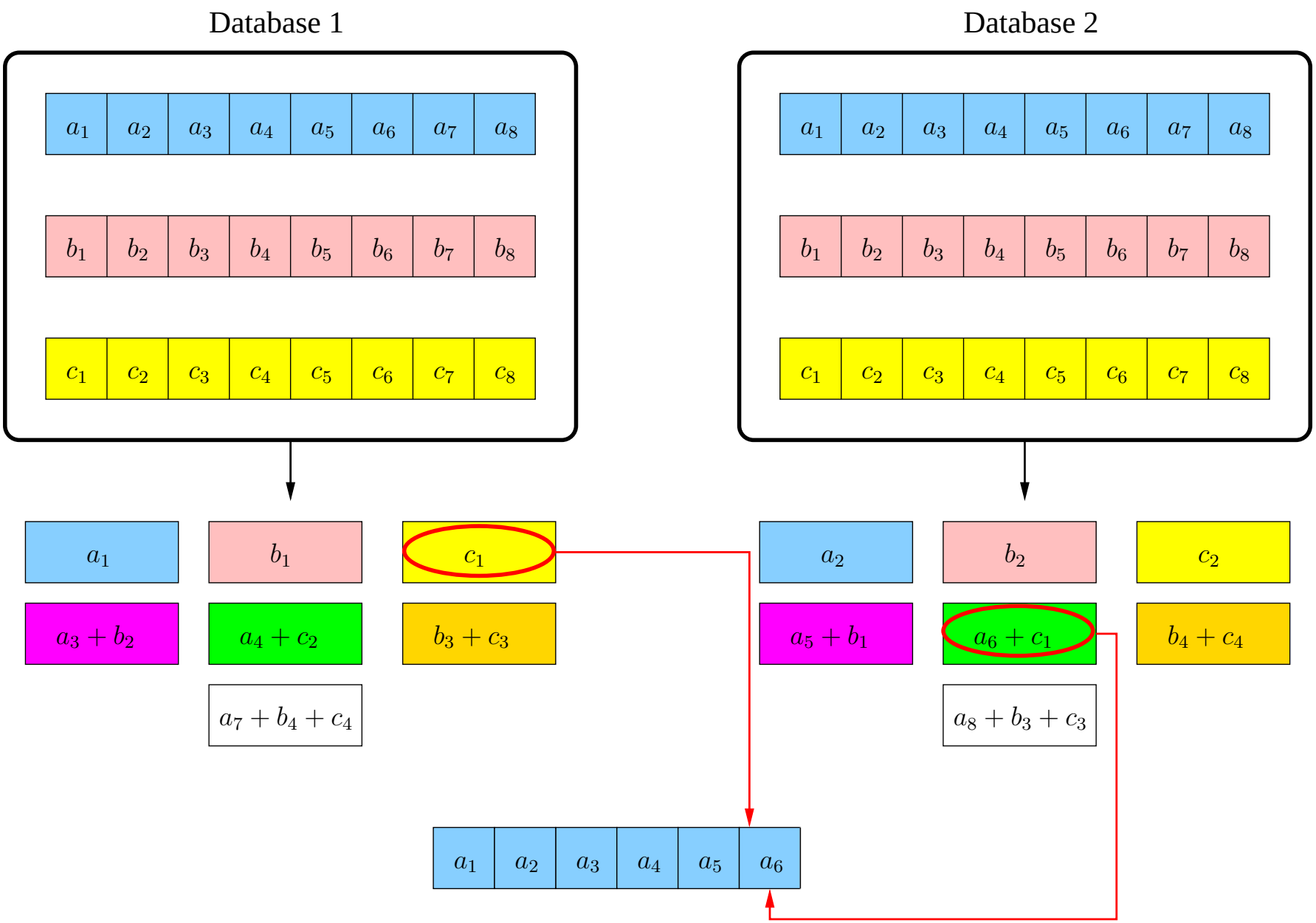
Example $M = 3, N = 2$: Decoding



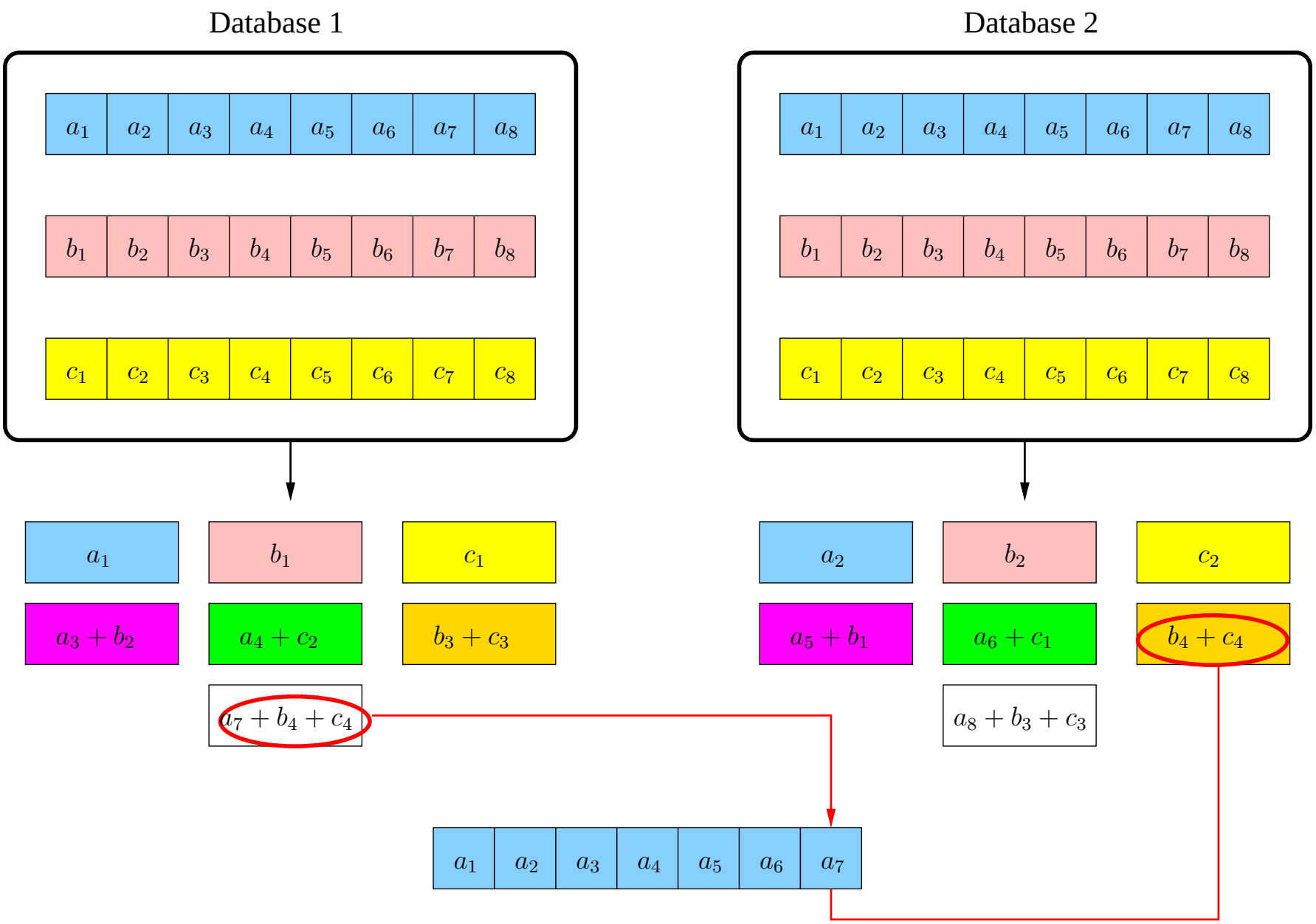
Example $M = 3, N = 2$: Decoding



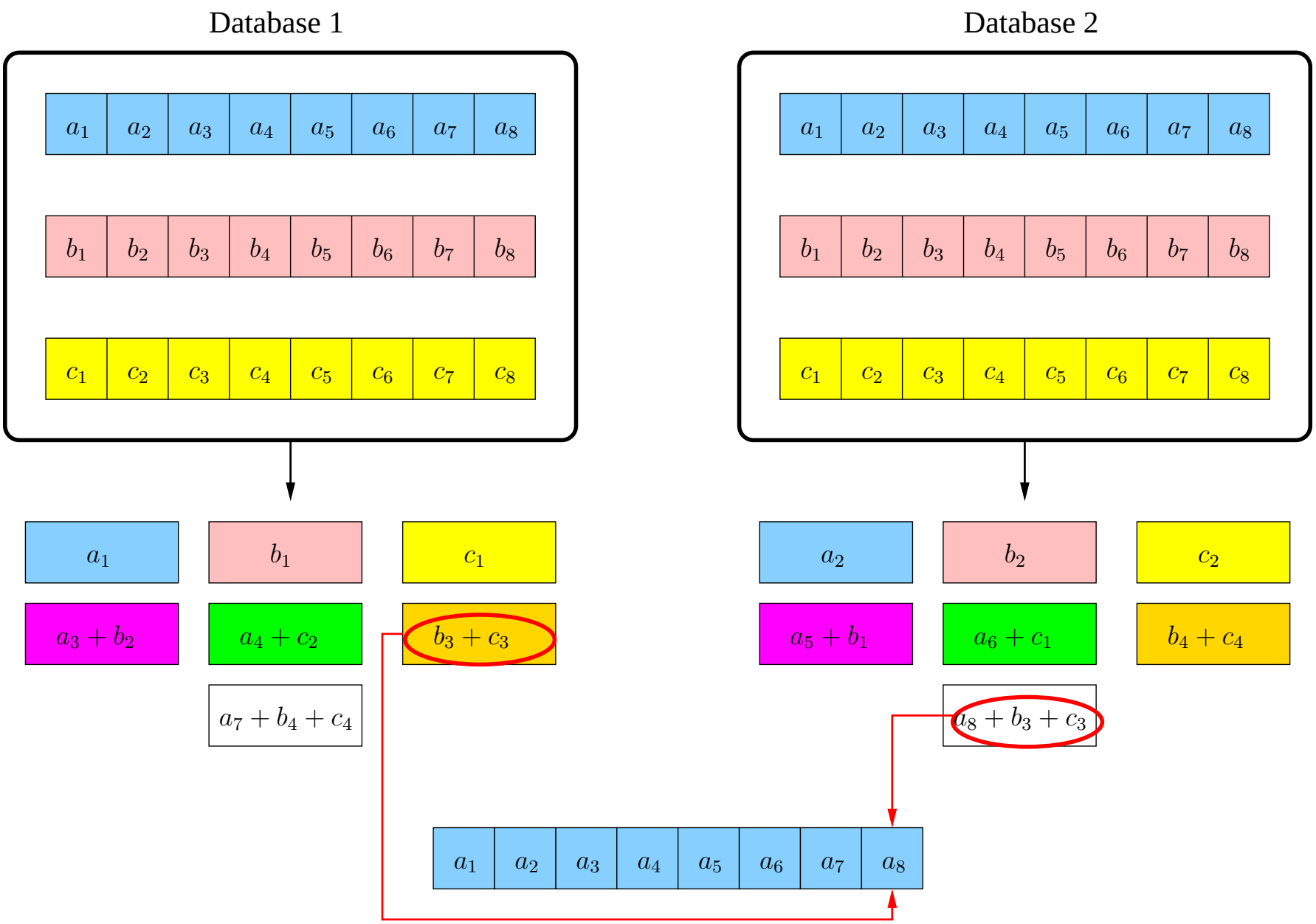
Example $M = 3, N = 2$: Decoding



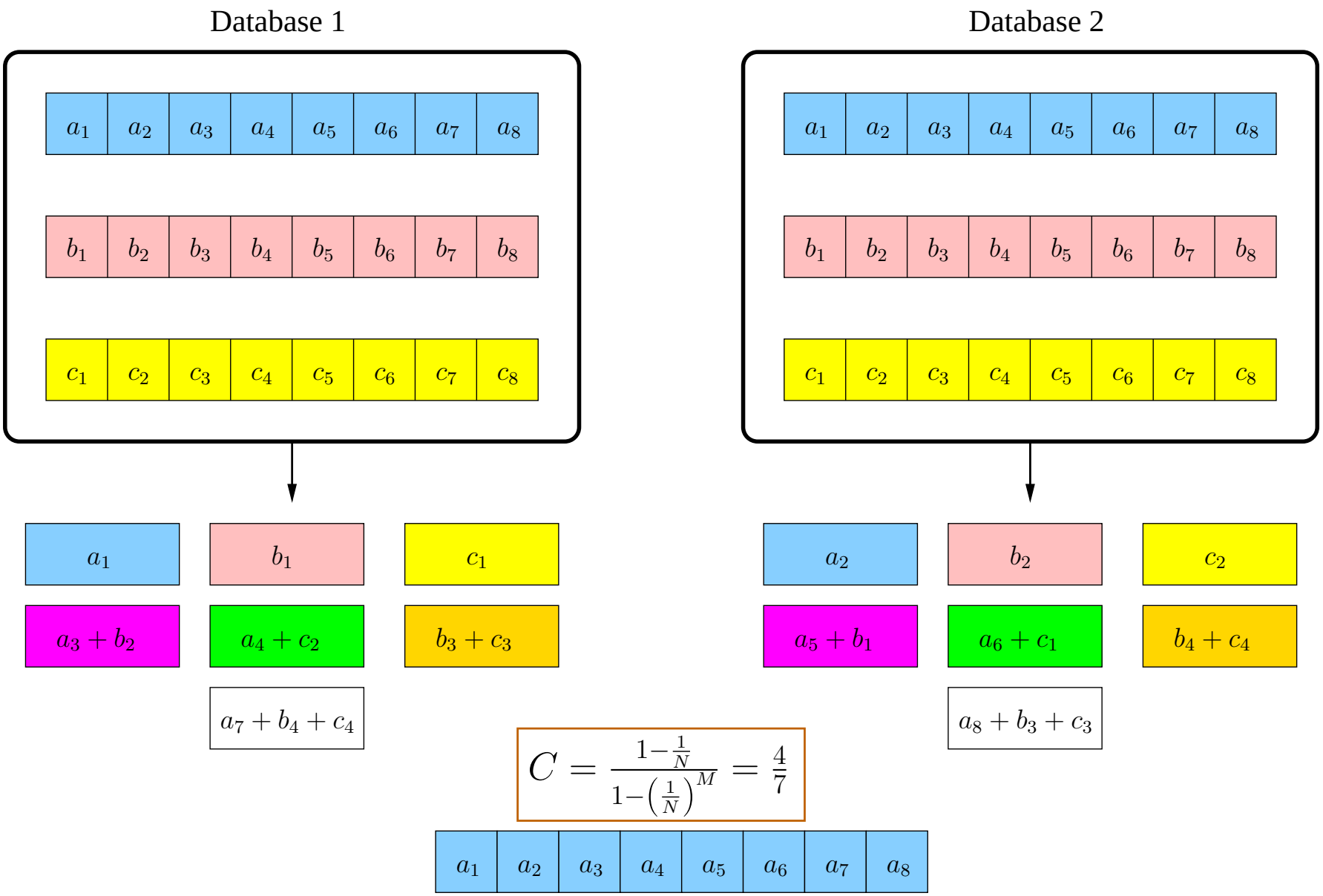
Example $M = 3, N = 2$: Decoding



Example $M = 3, N = 2$: Decoding



Example $M = 3, N = 2$: Private Retrieval Rate



Query Table for $M = 3, N = 2$

Database 1

a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8
b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8
c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8

Database 2

a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8
b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8
c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8

Database 1	Database 2
a_1 b_1 c_1	a_2 b_2 c_2
$a_3 + b_2$ $a_4 + c_2$ $b_3 + c_3$	$a_5 + b_1$ $a_6 + c_1$ $b_4 + c_4$
$a_7 + b_4 + c_4$	$a_8 + b_3 + c_3$

$$R = \frac{1 - \frac{1}{N}}{1 - (\frac{1}{N})^M} = \frac{1 - \frac{1}{2}}{1 - (\frac{1}{2})^3} = \frac{8}{14} = \frac{4}{7}$$

Query Table for $M = 2, N = 3$

Database 1	Database 2	Database 3
a_1 b_1	a_2 b_2	a_3 b_3
$a_4 + b_2$ $a_5 + b_3$	$a_6 + b_1$ $a_7 + b_3$	$a_8 + b_1$ $a_9 + b_2$

$$R = \frac{1 - \frac{1}{N}}{1 - (\frac{1}{N})^M} = \frac{1 - \frac{1}{3}}{1 - (\frac{1}{3})^2} = \frac{9}{12} = \frac{3}{4}$$

Query Table for $M = 3, N = 3$

Database 1	Database 2	Database 3
a_1	a_2	a_3
b_1	b_2	b_3
c_1	c_2	c_3
$a_4 + b_2$	$a_8 + b_1$	$a_{12} + b_1$
$a_5 + c_2$	$a_9 + c_1$	$a_{13} + c_1$
$b_4 + c_4$	$b_6 + c_6$	$b_8 + c_8$
$a_6 + b_3$	$a_{10} + b_3$	$a_{14} + b_2$
$a_7 + c_3$	$a_{11} + c_3$	$a_{15} + c_2$
$b_5 + c_5$	$b_7 + c_7$	$b_9 + c_9$
$a_{16} + b_6 + c_6$	$a_{20} + b_4 + c_4$	$a_{24} + b_4 + c_4$
$a_{17} + b_7 + c_7$	$a_{21} + b_5 + c_5$	$a_{25} + b_5 + c_5$
$a_{18} + b_8 + c_8$	$a_{22} + b_8 + c_8$	$a_{26} + b_6 + c_6$
$a_{19} + b_9 + c_9$	$a_{23} + b_9 + c_9$	$a_{27} + b_7 + c_7$

$$R = \frac{1 - \frac{1}{N}}{1 - (\frac{1}{N})^M} = \frac{1 - \frac{1}{3}}{1 - (\frac{1}{3})^2} = \frac{27}{39} = \frac{9}{13}$$

Interference Alignment in PIR

Database 1	Database 2	Database 3
a_1 b_1 c_1	a_2 b_2 c_2	a_3 b_3 c_3
$a_4 + b_2$ $a_5 + c_2$ $b_4 + c_4$ $a_6 + b_3$ $a_7 + c_3$ $b_5 + c_5$	$a_8 + b_1$ $a_9 + c_1$ $b_6 + c_6$ $a_{10} + b_3$ $a_{11} + c_3$ $b_7 + c_7$	$a_{12} + b_1$ $a_{13} + c_1$ $b_8 + c_8$ $a_{14} + b_2$ $a_{15} + c_2$ $b_9 + c_9$
$a_{16} + b_6 + c_6$ $a_{17} + b_7 + c_7$ $a_{18} + b_8 + c_8$ $a_{19} + b_9 + c_9$	$a_{20} + b_4 + c_4$ $a_{21} + b_5 + c_5$ $a_{22} + b_8 + c_8$ $a_{23} + b_9 + c_9$	$a_{24} + b_4 + c_4$ $a_{25} + b_5 + c_5$ $a_{26} + b_6 + c_6$ $a_{27} + b_7 + c_7$

$$R = \frac{1 - \frac{1}{N}}{1 - (\frac{1}{N})^M} = \frac{1 - \frac{1}{3}}{1 - (\frac{1}{3})^2} = \frac{27}{39} = \frac{9}{13}$$

Achievability Proof: Rate Calculation

- The scheme consists of M rounds.
- Each round uses side information from $N-1$ databases.
- **Number of stages:**
  - Round 1: 1 stage.
  - Round 2: $1 * (N-1)$ stages.
  - ...
  - Round k: $(N-1)^{k-1}$ stages.
- **Number of bits in each stage of the kth round:**
  - Total bits$= \binom{M}{k}$.
  - Desired bits$=\binom{M-1}{k-1}$.
- **Retrieval rate is given by:**

$$R = \frac{\text{Desired bits}}{\text{Total bits}} = \frac{\sum_{k=1}^{M} \binom{M-1}{k-1}(N-1)^{k-1}}{\sum_{k=1}^{M} \binom{M}{k}(N-1)^{k-1}}$$

$$= \frac{N^{M-1}}{\frac{1}{N-1}\sum_{k=1}^{M} \binom{M}{k}(N-1)^{k}}$$

$$= \frac{N^{M-1}}{\frac{1}{N-1}(N^M-1)}$$

$$= \frac{N^M - N^{M-1}}{N^M - 1} = \frac{1-\frac{1}{N}}{1-(\frac{1}{N})^M}$$

Query Table for $M = 3, N = 3$

		Database 1	Database 2	Database 3
round 1	stage 1	a_1 b_1 c_1	a_2 b_2 c_2	a_3 b_3 c_3
round 2	stage 1	$a_4 + b_2$ $a_5 + c_2$ $b_4 + c_4$	$a_8 + b_1$ $a_9 + c_1$ $b_6 + c_6$	$a_{12} + b_1$ $a_{13} + c_1$ $b_8 + c_8$
	stage 2	$a_6 + b_3$ $a_7 + c_3$ $b_5 + c_5$	$a_{10} + b_3$ $a_{11} + c_3$ $b_7 + c_7$	$a_{14} + b_2$ $a_{15} + c_2$ $b_9 + c_9$
round 3	stage 1	$a_{16} + b_6 + c_6$	$a_{20} + b_4 + c_4$	$a_{24} + b_4 + c_4$
	stage 2	$a_{17} + b_7 + c_7$	$a_{21} + b_5 + c_5$	$a_{25} + b_5 + c_5$
	stage 3	$a_{18} + b_8 + c_8$	$a_{22} + b_8 + c_8$	$a_{26} + b_6 + c_6$
	stage 4	$a_{19} + b_9 + c_9$	$a_{23} + b_9 + c_9$	$a_{27} + b_7 + c_7$

$$R = \frac{1 - \frac{1}{N}}{1 - (\frac{1}{N})^M} = \frac{1 - \frac{1}{3}}{1 - (\frac{1}{3})^2} = \frac{27}{39} = \frac{9}{13}$$

Converse Proof: Step 1

- **Lemma: Interference lower bound lemma**

$$\underbrace{\sum_{n=1}^{N} H(A_n^{[1]}) - L}_{\text{interference within answers}} + o(L) \geq \underbrace{I(W_{2:M}; Q_{1:N}^{[1]}, A_{1:N}^{[1]} | W_1)}_{\text{mutual information between interfering messages and the answers}}$$

- **Intuition**
  - In the absence of privacy constraint, user downloads L bits only.
  - The interference bits due to privacy is $\sum_{n=1}^{N} H(A_n^{[1]}) - L$.
  - In general, the answer strings are mixture of all messages.
  - Interference is no less than the mutual info. between $W_{2:M}$ and the answers.

Proof of Interference Lower Bound Lemma

$$
\begin{aligned}
&I(W_{2:M}; Q_{1:N}^{[1]}, A_{1:N}^{[1]} | W_1) \\
&\quad = I(W_{2:M}; Q_{1:N}^{[1]}, A_{1:N}^{[1]}, W_1) \quad \text{(independence of messages)} \\
&\quad = I(W_{2:M}; Q_{1:N}^{[1]}, A_{1:N}^{[1]}) + \underbrace{I(W_{2:M}; W_1 | Q_{1:N}^{[1]}, A_{1:N}^{[1]})}_{=o(L)\text{(reliability)}} \\
&\quad = I(W_{2:M}; A_{1:N}^{[1]} | Q_{1:N}^{[1]}) + o(L) \quad \text{(independence of queries and messages)} \\
&\quad = \underbrace{H(A_{1:N}^{[1]} | Q_{1:N}^{[1]})}_{\leq \sum_{n=1}^{N} H(A_n^{[1]})} - \underbrace{H(A_{1:N}^{[1]} | Q_{1:N}^{[1]}, W_{2:M})}_{\text{include } W_1 \text{ back}} + o(L) \\
&\quad \leq \sum_{n=1}^{N} H(A_n^{[1]}) - H(A_{1:N}^{[1]}, W_1 | Q_{1:N}^{[1]}, W_{2:M}) + \underbrace{H(W_1 | A_{1:N}^{[1]}, Q_{1:N}^{[1]}, W_{2:M})}_{=o(L)\text{(reliability)}} + o(L) \\
&\quad = \sum_{n=1}^{N} H(A_n^{[1]}) - \underbrace{H(W_1 | Q_{1:N}^{[1]}, W_{2:M})}_{=L \text{ (independence)}} - \underbrace{H(A_{1:N}^{[1]} | Q_{1:N}^{[1]}, W_{1:M})}_{=0 \text{ (answers are functions of msgs and queries)}} + o(L) \\
&\quad = \sum_{n=1}^{N} H(A_n^{[1]}) - L + o(L)
\end{aligned}
$$

Converse Proof: Step 2

- **Lemma: Induction lemma**

$$I(W_{m:M}; Q_{1:N}^{[m-1]}, A_{1:N}^{[m-1]} | W_{1:m-1}) \geq \frac{1}{N} I(W_{m+1:M}; Q_{1:N}^{[m]}, A_{1:N}^{[m]} | W_{1:m}) + \frac{L}{N} - \frac{o(L)}{N}$$

- **Intuition**
  - Relates the interference bounds if the number of messages decreases by 1.
  - Constructs an induction relation for the interference bounds.
  - Consequence of the privacy constraint.

Proof of Induction Lemma

$$
\begin{aligned}
& I(W_{m:M}; Q_{1:N}^{[m-1]}, A_{1:N}^{[m-1]} | W_{1:m-1}) \\
& \geq \frac{1}{N} \sum_{n=1}^{N} I(W_{m:M}; Q_n^{[m-1]}, A_n^{[m-1]} | W_{1:m-1}) \quad \text{(mutual info. is non-negative)} \\
& = \frac{1}{N} \sum_{n=1}^{N} I(W_{m:M}; Q_n{}^{[m]}, A_n^{[m]} | W_{1:m-1}) \quad \text{(privacy)} \\
& = \frac{1}{N} \sum_{n=1}^{N} I(W_{m:M}; A_n^{[m]} | Q_n^{[m]}, W_{1:m-1}) \quad \text{(independence of msgs and queries)} \\
& = \frac{1}{N} \sum_{n=1}^{N} H(A_n^{[m]} | Q_n^{[m]}, W_{1:m-1}) - \underbrace{H(A_n^{[m]} | Q_n^{[m]}, W_{1:M})}_{\substack{=0 \text{ (answers are functions} \\ \text{of msgs and queries)}}} \\
& \geq \frac{1}{N} \sum_{n=1}^{N} H(A_n^{[m]} | Q_{1:N}^{[m]}, A_{1:n-1}^{[m]}, W_{1:m-1}) \quad \text{(conditioning reduces entropy)} \\
& = \frac{1}{N} \sum_{n=1}^{N} H(A_n^{[m]} | Q_{1:N}^{[m]}, A_{1:n-1}^{[m]}, W_{1:m-1}) - \underbrace{H(A_n^{[m]} | Q_{1:N}^{[m]}, A_{1:n-1}^{[m]}, W_{1:M})}_{\substack{=0 \text{ (answers are functions} \\ \text{of msgs and queries)}}}
\end{aligned}
$$

Proof of Induction Lemma (cont.)

$$
\begin{aligned}
& I(W_{m:M}; Q_{1:N}^{[m-1]}, A_{1:N}^{[m-1]} | W_{1:m-1}) \\
& \geq \frac{1}{N} \sum_{n=1}^{N} H(A_n^{[m]} | Q_{1:N}^{[m]}, A_{1:n-1}^{[m]}, W_{1:m-1}) - H(A_n^{[m]} | Q_{1:N}^{[m]}, A_{1:n-1}^{[m]}, W_{1:M}) \\
& = \frac{1}{N} \sum_{n=1}^{N} I(W_{m:M}; A_n^{[m]} | Q_{1:N}^{[m]}, A_{1:n-1}^{[m]}, W_{1:m-1}) \\
& = \frac{1}{N} I(W_{m:M}; A_{1:N}^{[m]} | Q_{1:N}^{[m]}, W_{1:m-1}) \quad \text{(chain rule)} \\
& = \frac{1}{N} I(W_{m:M}; Q_{1:N}^{[m]}, A_{1:N}^{[m]} | W_{1:m-1}) \quad \text{(independence of msgs and queries)} \\
& = \frac{1}{N} I(W_{m:M}; W_m, Q_{1:N}^{[m]}, A_{1:N}^{[m]} | W_{1:m-1}) - \frac{1}{N} \underbrace{I(W_{m:M}; W_m | W_{1:m-1}, Q_{1:N}^{[m]}, A_{1:N}^{[m]})}_{=o(L) \text{ (reliability)}} \\
& = \frac{1}{N} \underbrace{I(W_{m:M}; W_m | W_{1:m-1})}_{=L \text{ (independence)}} + \frac{1}{N} I(\underbrace{W_{m:M}}_{\text{drop } W_m}; Q_{1:N}^{[m]}, A_{1:N}^{[m]} | W_{1:m}) - \frac{o(L)}{N} \\
& = \frac{L}{N} + \frac{1}{N} I(W_{m+1:M}; Q_{1:N}^{[m]}, A_{1:N}^{[m]} | W_{1:m}) - \frac{o(L)}{N}
\end{aligned}
$$

Proof of Induction Lemma (cont.)

$$
\begin{aligned}
& I(W_{m:M}; Q_{1:N}^{[m-1]}, A_{1:N}^{[m-1]} | W_{1:m-1}) \\
& \geq \frac{1}{N} \sum_{n=1}^{N} H(A_n^{[m]} | Q_{1:N}^{[m]}, A_{1:n-1}^{[m]}, W_{1:m-1}) - H(A_n^{[m]} | Q_{1:N}^{[m]}, A_{1:n-1}^{[m]}, W_{1:M}) \\
& = \frac{1}{N} \sum_{n=1}^{N} I(W_{m:M}; A_n^{[m]} | Q_{1:N}^{[m]}, A_{1:n-1}^{[m]}, W_{1:m-1}) \\
& = \frac{1}{N} I(W_{m:M}; A_{1:N}^{[m]} | Q_{1:N}^{[m]}, W_{1:m-1}) \quad \text{(chain rule)} \\
& = \frac{1}{N} I(W_{m:M}; Q_{1:N}^{[m]}, A_{1:N}^{[m]} | W_{1:m-1}) \quad \text{(independence of msgs and queries)} \\
& = \frac{1}{N} I(W_{m:M}; W_m, Q_{1:N}^{[m]}, A_{1:N}^{[m]} | W_{1:m-1}) - \frac{1}{N} \underbrace{I(W_{m:M}; W_m | W_{1:m-1}, Q_{1:N}^{[m]}, A_{1:N}^{[m]})}_{=o(L) \text{ (reliability)}} \\
& = \frac{1}{N} \underbrace{I(W_{m:M}; W_m | W_{1:m-1})}_{=L \text{ (independence)}} + \frac{1}{N} I(\underbrace{W_{m:M}}_{\text{drop } W_m}; Q_{1:N}^{[m]}, A_{1:N}^{[m]} | W_{1:m}) - \frac{o(L)}{N} \\
& = \frac{L}{N} + \frac{1}{N} I(W_{m+1:M}; Q_{1:N}^{[m]}, A_{1:N}^{[m]} | W_{1:m}) - \frac{o(L)}{N}
\end{aligned}
$$

Converse Proof: Fundamental Lemmas

- **Lemma: Interference lower bound lemma**

$$\sum_{n=1}^{N} H(A_n^{[1]}) - L + o(L) \geq I(W_{2:M}; Q_{1:N}^{[1]}, A_{1:N}^{[1]} | W_1)$$

- **Lemma: Induction lemma**

$$I(W_{m:M}; Q_{1:N}^{[m-1]}, A_{1:N}^{[m-1]} | W_{1:m-1})$$
$$\geq \frac{1}{N} I(W_{m+1:M}; Q_{1:N}^{[m]}, A_{1:N}^{[m]} | W_{1:m}) + \frac{L}{N} - \frac{o(L)}{N}$$

Converse Proof: Main Body

- **Applying Interference lower bound and induction lemmas**

$$\sum_{n=1}^{N} H(A_n^{[1]}) - L + o(L)$$

$$\geq I(W_{2:M}; Q_{1:N}^{[1]}, A_{1:N}^{[1]} | W_1) \quad \text{(interference lower bound lemma)}$$

$$\geq \frac{1}{N} I(W_{3:M}; Q_{1:N}^{[2]}, A_{1:N}^{[2]} | W_{1:2}) + \frac{L}{N} - \frac{o(L)}{N} \quad \text{(induction lemma)}$$

$$\geq \frac{1}{N}\left[I(W_{4:M}; Q_{1:N}^{[3]}, A_{1:N}^{[3]} | W_{1:3}) + \frac{L}{N} - \frac{o(L)}{N}\right] + \frac{L}{N} - \frac{o(L)}{N} \quad \text{(induction lemma)}$$

$$\geq \cdots \text{(induction lemma} \cdots \text{ induction lemma} \cdots \text{ induction lemma} \cdots)$$

$$\geq \left(\frac{1}{N} + \frac{1}{N^2} + \cdots + \frac{1}{N^{M-1}}\right)(L - o(L))$$

- **Reordering terms**

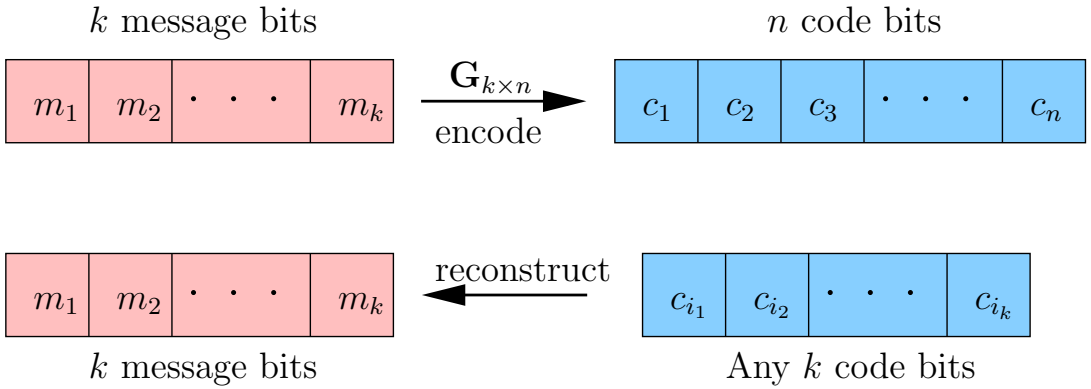
$$\sum_{n=1}^{N} H(A_n^{[1]}) \geq \left(1 + \frac{1}{N} + \frac{1}{N^2} + \cdots + \frac{1}{N^{M-1}}\right)(L - o(L))$$

- **Dividing by L and taking $L \to \infty$**

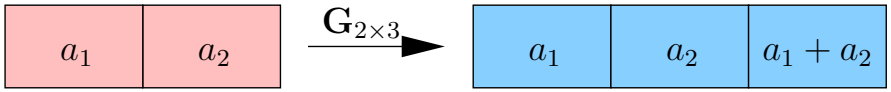
$$R = \frac{L}{\sum_{n=1}^{N} H(A_n^{[1]})} \leq \left(1 + \frac{1}{N} + \frac{1}{N^2} + \cdots + \frac{1}{N^{M-1}}\right)^{-1} = \frac{1 - \frac{1}{N}}{1 - (\frac{1}{N})^M}$$

(n, k) MDS Code

- Construct the n-length codeword by observing any k coded symbols.



- Any $k \times k$ submatrix $\mathbf{G}_{k\times k}$ of its generator matrix $\mathbf{G}$ is full rank.
- **Example: (3,2) MDS code**

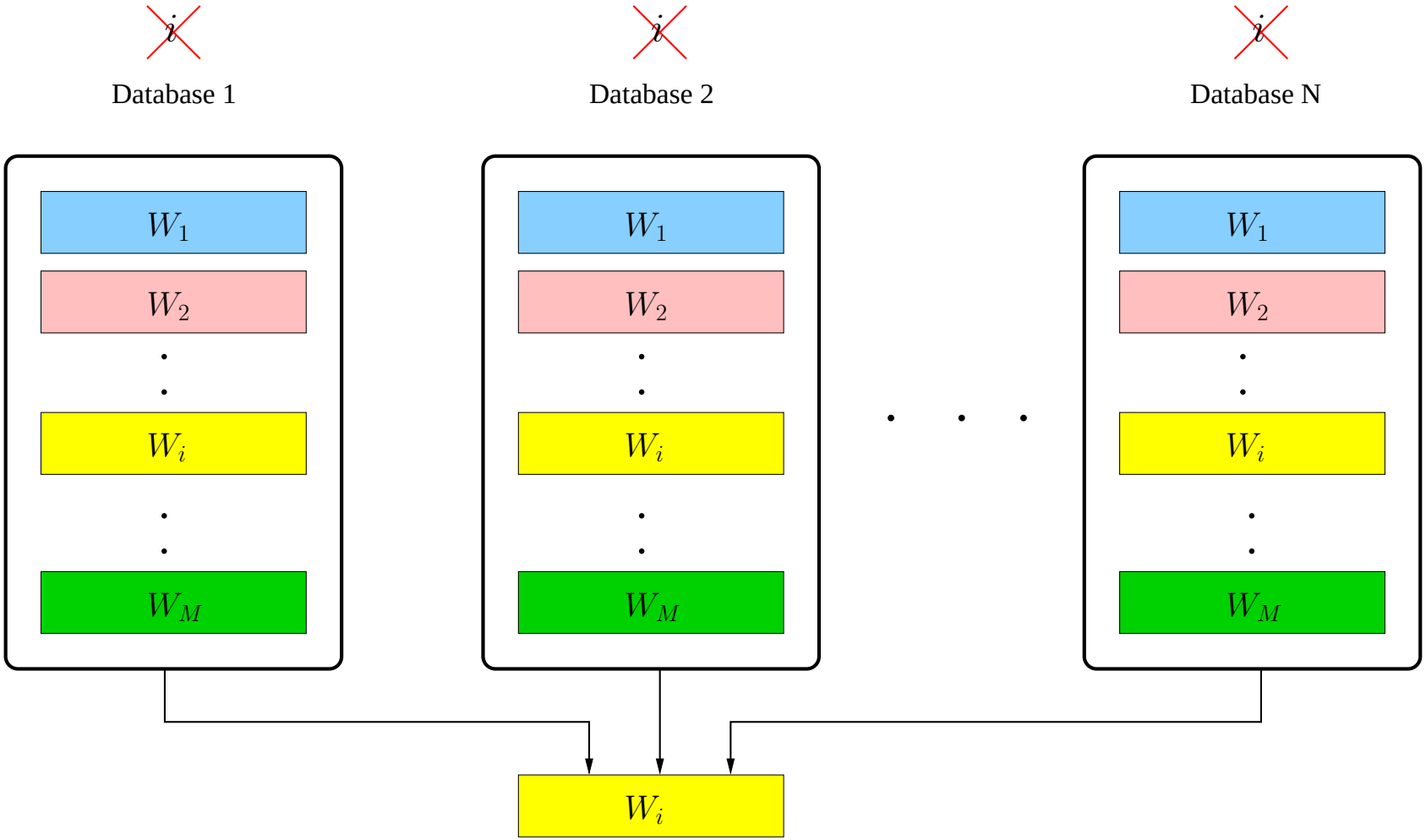


$$\mathbf{G}_{2\times 3} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

- Maximum distance separable codes: achieves Singleton bound.

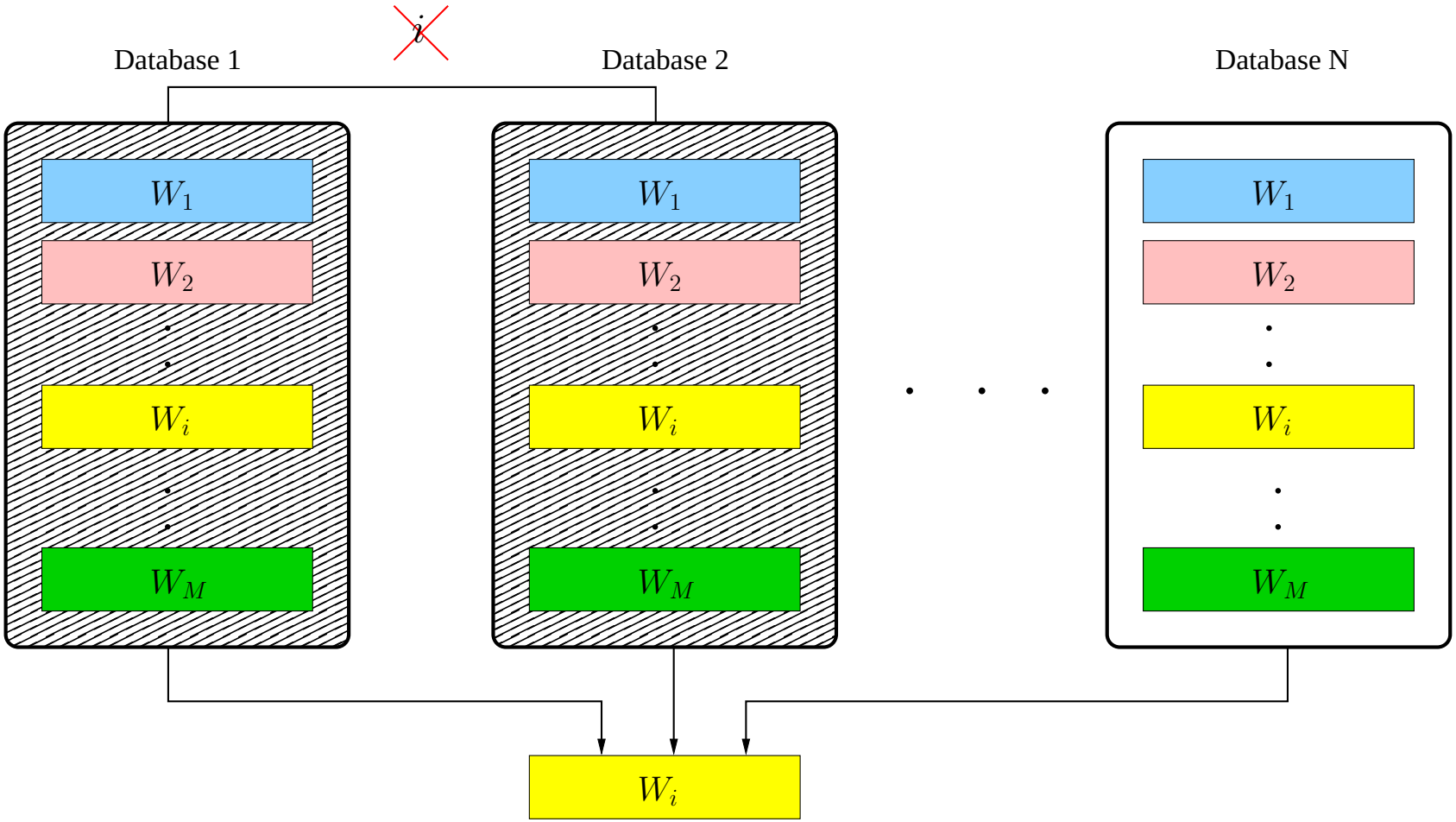
$$d = n - k + 1$$

PIR with Colluding Databases (TPIR) [Sun-Jafar][4]



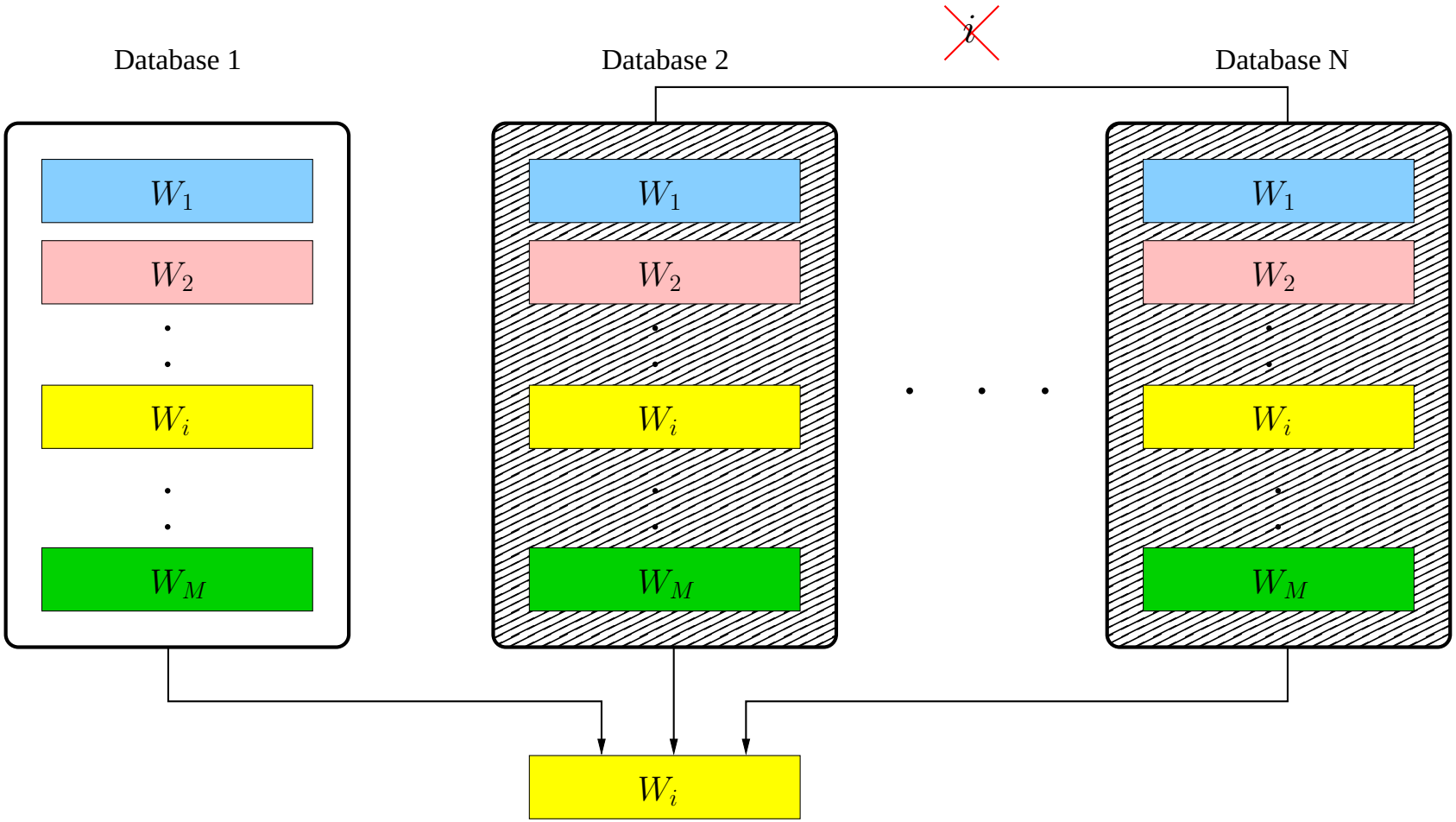
[4] H. Sun and S. A. Jafar. The capacity of robust private information retrieval with colluding databases. IEEE Trans. on Info. Theory, 64(4):2361–2370, April 2018.

PIR with Colluding Databases (TPIR) [Sun-Jafar][4]



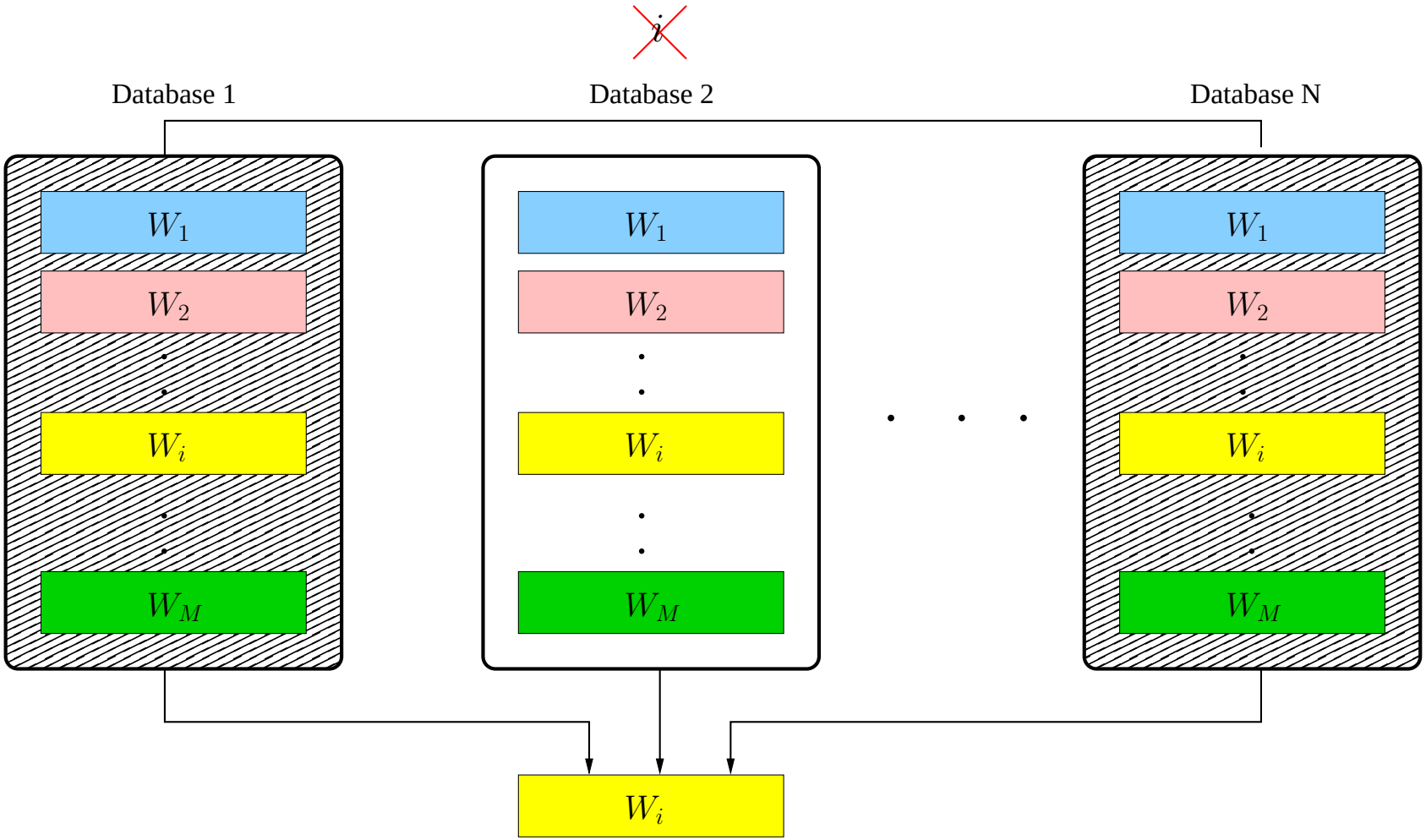
[4] H. Sun and S. A. Jafar. The capacity of robust private information retrieval with colluding databases. IEEE Trans. on Info. Theory, 64(4):2361–2370, April 2018.

PIR with Colluding Databases (TPIR) [Sun-Jafar][4]



[4] H. Sun and S. A. Jafar. The capacity of robust private information retrieval with colluding databases. IEEE Trans. on Info. Theory, 64(4):2361–2370, April 2018.

PIR with Colluding Databases (TPIR) [Sun-Jafar][4]



[4] H. Sun and S. A. Jafar. The capacity of robust private information retrieval with colluding databases. IEEE Trans. on Info. Theory, 64(4):2361–2370, April 2018.

Back to the Query Table for $M = 2, N = 3$

Database 1	Database 2	Database 3
a_1 b_1	a_2 b_2	a_3 b_3
$a_4 + b_2$ $a_5 + b_3$	$a_6 + b_1$ $a_7 + b_3$	$a_8 + b_1$ $a_9 + b_2$

Back to the Query Table for $M = 2, N = 3$: $T = 2$ Collude

Database 1	Database 2	Database 3
a_1 b_1	a_2 b_2	a_3 b_3
$a_4 + b_2$ $a_5 + b_3$	$a_6 + b_1$ $a_7 + b_3$	$a_8 + b_1$ $a_9 + b_2$

Back to the Query Table for $M = 2, N = 3$: $T = 2$ Collude

Database 1	Database 2	Database 3
a_1 b_1	a_2 b_2	a_3 b_3
$a_4 + b_2$ $a_5 + b_3$	$a_6 + b_1$ $a_7 + b_3$	$a_8 + b_1$ $a_9 + b_2$

Back to the Query Table for $M = 2, N = 3$: $T = 2$ Collude

Database 1	Database 2	Database 3
a_1 b_1	a_2 b_2	a_3 b_3
$a_4 + b_2$ $a_5 + b_3$	$a_6 + b_1$ $a_7 + b_3$	$a_8 + b_1$ $a_9 + b_2$

- **Privacy is compromised.**

TPIR Example: $M = 2, N = 3, T = 2$

Database 1	Database 2	Database 3
a_1 a_2 b_1 b_2	a_3 a_4 b_3 b_4	a_5 a_6 b_5 b_6
$a_7 + b_7$	$a_8 + b_8$	$a_9 + b_9$

$$a_{[1:9]} = \mathbf{S}_{1_{9\times 9}} W_1$$
$$b_{[1:9]} = \mathrm{MDS}_{9\times 6} \mathbf{S}_2([1:6], :)_{6\times 9} W_2$$

TPIR Example: $M = 2, N = 3, T = 2$

Database 1	Database 2	Database 3
a_1 a_2 b_1 b_2	a_3 a_4 b_3 b_4	a_5 a_6 b_5 b_6
$a_7 + b_7$	$a_8 + b_8$	$a_9 + b_9$

$$a_{[1:9]} = \mathbf{S}_{1_{9\times 9}} W_1$$
$$b_{[1:9]} = \mathrm{MDS}_{9\times 6} \mathbf{S}_2([1:6], :)_{6\times 9} W_2$$

- Download randomly-mixed symbols from desired and undesired messages.
- The undesired messages are further encoded by MDS code.

TPIR Example: $M = 2, N = 3, T = 2$

Database 1	Database 2	Database 3
a_1	a_3	a_5
a_2	a_4	a_6
b_1	b_3	b_5
b_2	b_4	b_6
$a_7 + \cancel{b_7}$	$a_8 + \cancel{b_8}$	$a_9 + \cancel{b_9}$

$$a_{[1:9]} = \mathbf{S}_{1_{9\times 9}} W_1$$

$$b_{[1:9]} = \mathrm{MDS}_{9\times 6} \mathbf{S}_2([1:6], :)_{6\times 9} W_2$$

TPIR Example: $M = 2, N = 3, T = 2$

Database 1	Database 2	Database 3
a_1 a_2 b_1 b_2	a_3 a_4 b_3 b_4	a_5 a_6 b_5 b_6
a_7	a_8	a_9

$$a_{[1:9]} = \mathbf{S}_{1_{9\times 9}} W_1$$
$$b_{[1:9]} = \mathrm{MDS}_{9\times 6} \mathbf{S}_2([1:6], :)_{6\times 9} W_2$$

$$\boxed{R = \frac{1 - \frac{T}{N}}{1 - (\frac{T}{N})^M} = \frac{1 - \frac{2}{3}}{1 - (\frac{2}{3})^2} = \frac{9}{15} = \frac{3}{5}}$$

TPIR Example: $M = 2, N = 3, T = 2$

Database 1	Database 2	Database 3
a_1 a_2 b_1 b_2	a_3 a_4 b_3 b_4	a_5 a_6 b_5 b_6
$a_7 + b_7$	$a_8 + b_8$	$a_9 + b_9$

- Privacy: From any 2 databases

$$
\begin{aligned}
a_{\mathcal{I}} &= \mathbf{S}_1(\mathcal{I}, :) W_1 \\
&\sim \mathbf{S}_1([1:6], :) W_1 \\
b_{\mathcal{I}} &= MDS(\mathcal{I}, :)_{6\times 6} \mathbf{S}_2([1:6], :) W_2 \\
&\sim \mathbf{S}_2([1:6], :) W_2
\end{aligned}
$$

On the PIR Capacity Formula

- Capacity of classical PIR:

$$C_{\text{PIR}} = \frac{1 - \frac{1}{N}}{1 - (\frac{1}{N})^M}$$

  - Monotonically increasing in N.
  - Monotonically decreasing in M.

- Capacity of T-colluding PIR:

$$C_{\text{COL}} = \frac{1 - \frac{T}{N}}{1 - (\frac{T}{N})^M}$$

- **Effect of colluding:** divide N by T.

$$C_{\text{COL}} < C_{\text{PIR}}$$

Converse Proof: New Tools and Lemmas

- **Changes in the model:**
  - Privacy constraint: For any $\mathcal{T} \subset \{1, \cdots, N\}$ such that $|\mathcal{T}| = T$.

$$I(i; Q_{\mathcal{T}}^{[i]}) = 0$$

- Use Han's inequality

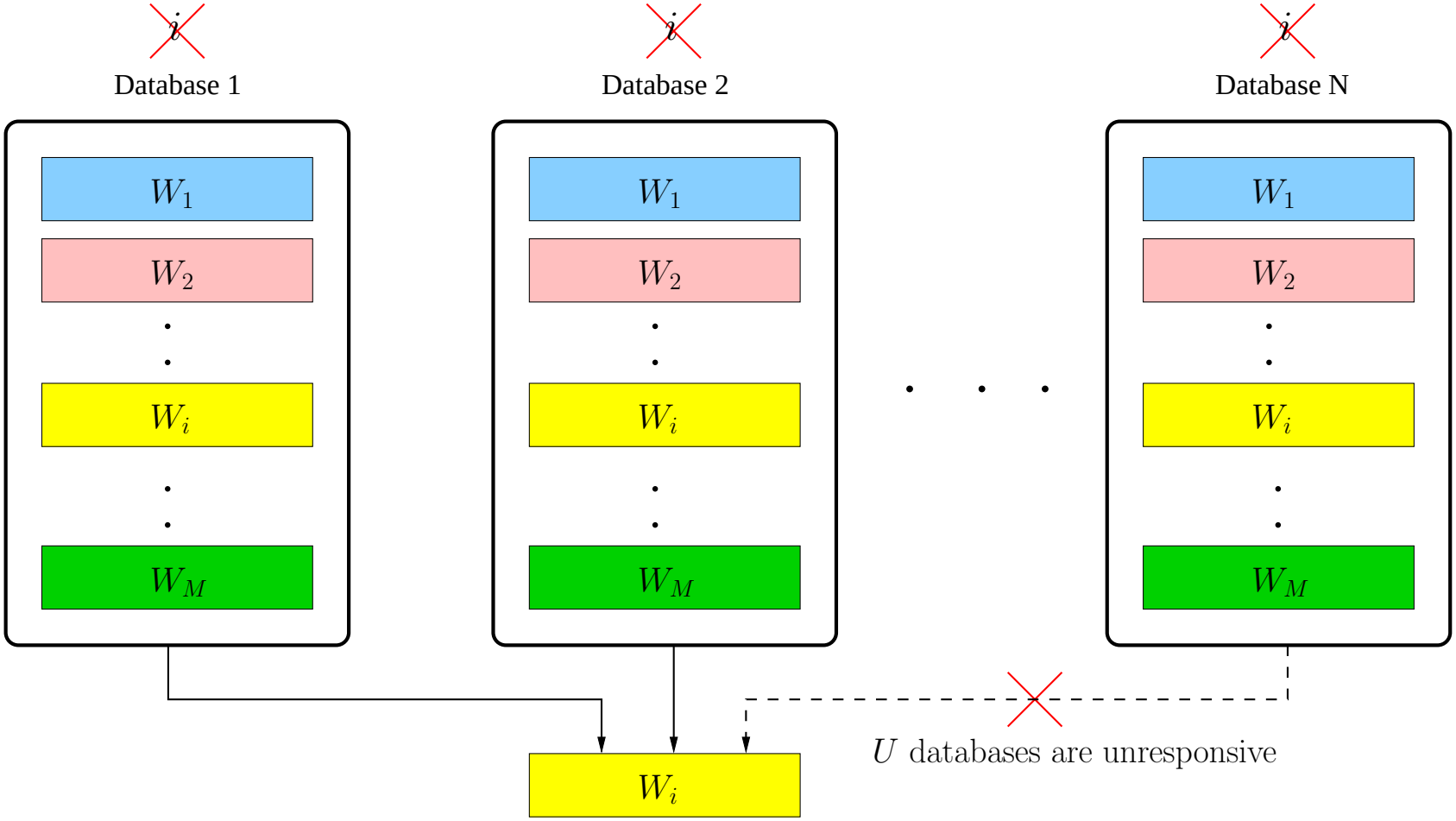
$$\frac{1}{\binom{N}{T}} \sum_{\mathcal{T}:|\mathcal{T}|=T} H(A_{\mathcal{T}}^{[m]} | Q_{1:N}^{[m]}, W_{1:m-1}) \geq \frac{T}{N} H(A_{1:N}^{[1]} | Q_{1:N}^{[m]}, W_{1:m-1})$$

- **Lemma: Interference lower bound lemma**
  - No change as it is not a consequence of privacy constraint.

- **Lemma: Induction lemma**

$$I(W_{m:M}; Q_{1:N}^{[m-1]}, A_{1:N}^{[m-1]} | W_{1:m-1})$$
$$\geq \frac{T}{N} I(W_{m+1:M}; Q_{1:N}^{[m]}, A_{1:N}^{[m]} | W_{1:m}) + \frac{TL}{N} - \frac{o(L)}{N}$$

Robust PIR (RPIR) [Sun-Jafar][4]



[4] H. Sun and S. A. Jafar. The capacity of robust private information retrieval with colluding databases. IEEE Trans. on Info. Theory, 64(4):2361–2370, April 2018.

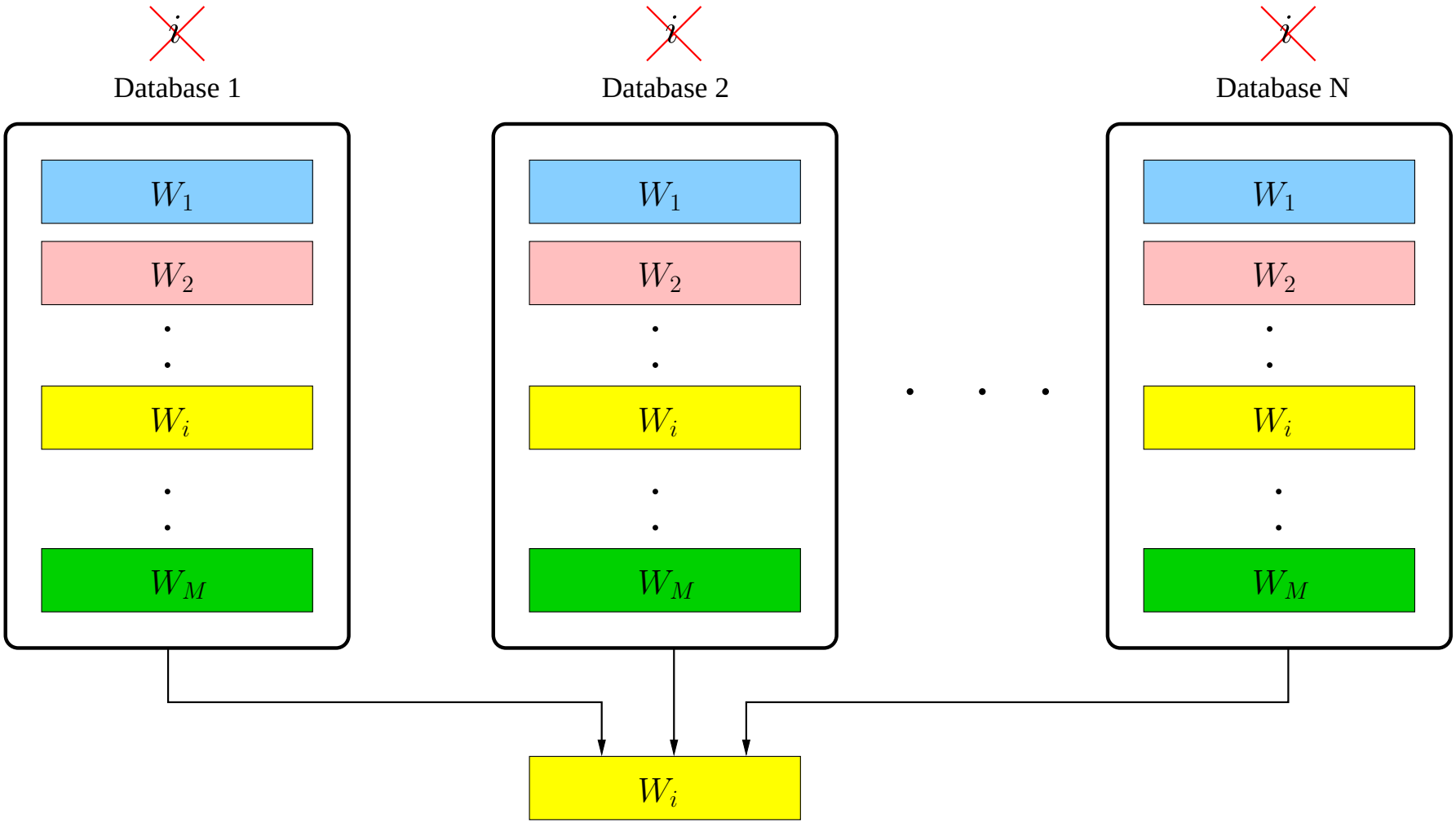
Achievable Scheme and Main Result

- **Modification to TPIR scheme:**
  - The unresponsive databases introduce erasures.
  - Use erasure code: $(\frac{N}{N-U}L, L)$ MDS code for the desired message.
  - Download MDS-coded mixtures from both desired and undesired messages.
- **Main result for robust-colluding PIR:**

$$\boxed{C = \frac{1 - \frac{T}{N-U}}{1 - (\frac{T}{N-U})^M}}$$

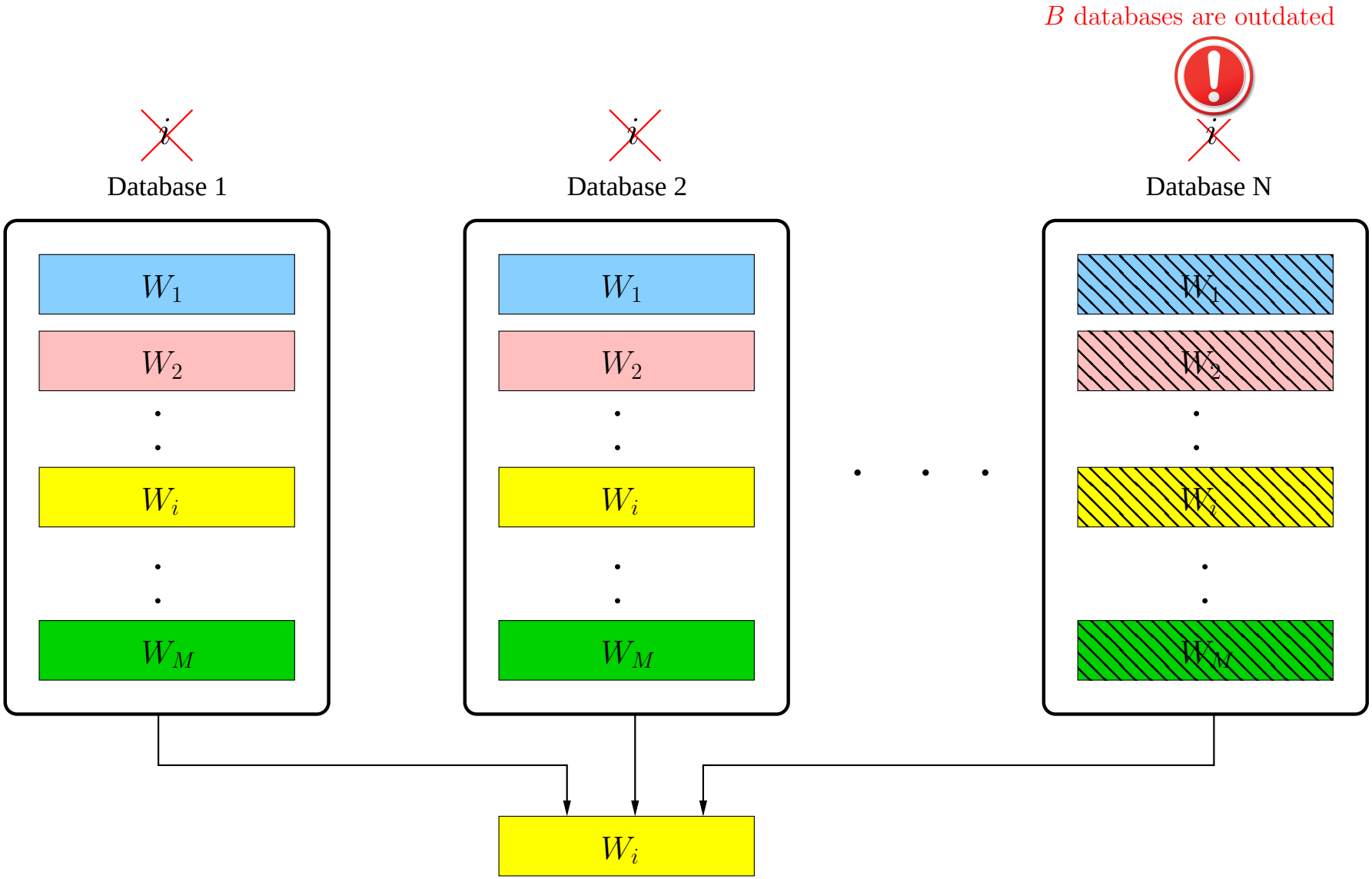
- **Converse:**
  - Only $N - U$ databases respond with answer strings.
  - Capacity cannot be larger than the capacity of TPIR with $N - U$ databases.
- No capacity penalty from not knowing which databases will respond.

PIR from Byzantine and Colluding Databases (BPIR) [Banawan-Ulukus][5]



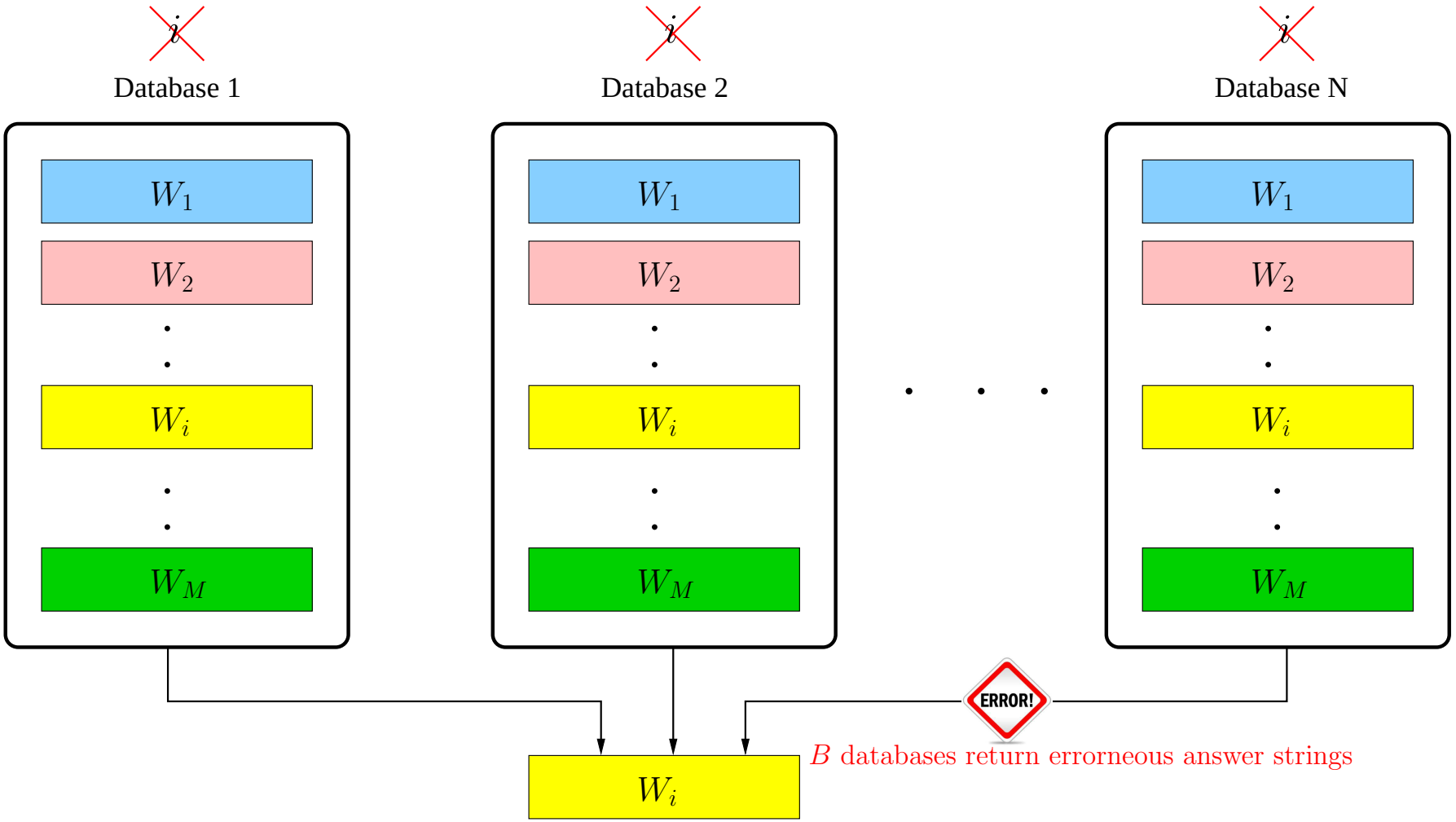
[5] K. Banawan and S. Ulukus, "The Capacity of Private Information retrieval from Byzantine and Colluding Databases," IEEE Trans. on Information Theory, submitted June 2017. Available on arXiv:1706.01442.

PIR from Byzantine and Colluding Databases (BPIR) [Banawan-Ulukus][5]



[5] K. Banawan and S. Ulukus, "The Capacity of Private Information retrieval from Byzantine and Colluding Databases," IEEE Trans. on Information Theory, submitted June 2017. Available on arXiv:1706.01442.

PIR from Byzantine and Colluding Databases (BPIR) [Banawan-Ulukus][5]



[5] K. Banawan and S. Ulukus, "The Capacity of Private Information retrieval from Byzantine and Colluding Databases," IEEE Trans. on Information Theory, submitted June 2017. Available on arXiv:1706.01442.

Achievable Scheme and Main Result

- **Modification to RPIR scheme:**
  - The Byzantine databases introduce errors.
  - For the undesired messages:
    - Encoded via punctured MDS code at every round.
    - Successive interference cancellation of side information.
  - The desired message is encoded by an outer $(\frac{N}{N-2B}L, L)$ MDS code.
- **Main result for BPIR:**

$$\boxed{C = \frac{N-2B}{N} \cdot \frac{1-\frac{T}{N-2B}}{1-(\frac{T}{N-2B})^M}}$$

- **Compared to TPIR result:**
  - Harm is equivalent to removing $2B$ storage nodes.
  - Penalty term $\frac{N-2B}{N}$: user needs to download from all N databases.

Comparing Capacity Formulas

- Classical PIR:

$$C_{\text{PIR}} = \frac{1 - \frac{1}{N}}{1 - (\frac{1}{N})^M}$$

- T-colluding PIR:

$$C_{\text{COL}} = \frac{1 - \frac{T}{N}}{1 - (\frac{T}{N})^M}$$

- U-robust PIR:

$$C_{\text{ROB}} = \frac{1 - \frac{T}{N-U}}{1 - (\frac{T}{N-U})^M}$$

- B-Byzantine PIR:

$$C_{\text{BYZ}} = \frac{N-2B}{N} \cdot \frac{1 - \frac{T}{N-2B}}{1 - (\frac{T}{N-2B})^M}$$

Converse Proof: Assumptions and Basic Tools

- **Byzantine databases are restricted to altering the contents.**
  - The nth Byzantine database changes its contents Ω_n from $\mathcal{W}$ to $\tilde{\mathcal{W}}$.
  - Weaker adversary $\Rightarrow$ potentially higher rate.
- **Answer is a deterministic function.**
  - $A_n^{[i]} = f_n(\Omega_n, Q_n^{[i]}) = A_n^{[i]}(\Omega_n)$, of the altered database (if Byzantine) Ω_n.
  - Weaker adversary $\Rightarrow$ potentially higher rate.
- **Retrieval scheme is symmetric.**
  - Any asymmetric scheme can be made symmetric by proper time-sharing.

Converse Proof: New Tools and Lemmas

- **Lemma:** **Uniqueness lemma**

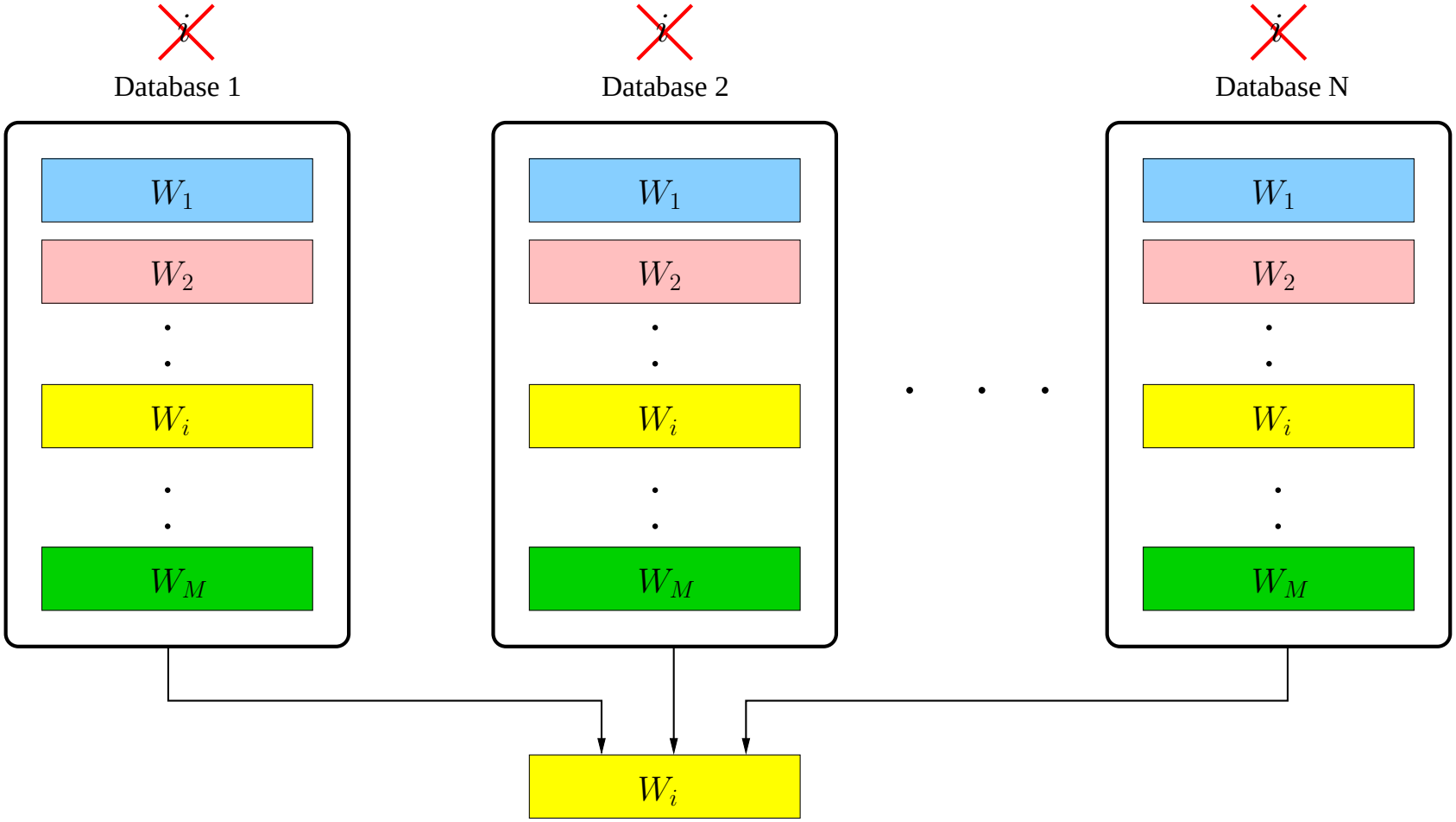
  Fix a set of honest databases $\mathcal{U} \subset \{1, \cdots, N\}$ such that $|\mathcal{U}| = N - 2B$, and $\Omega_n = \mathcal{W}$, for every $n \in \mathcal{U}$. Then, for correct decoding of W_i, the answer strings $A_{\mathcal{U}}^{[i]}(\mathcal{W})$ is unique for every realization of $\mathcal{W}$.

- **Intuition**
  - For different realizations of messages $\mathcal{W} \neq \tilde{\mathcal{W}}$, we have $A_{\mathcal{U}}^{[i]}(\mathcal{W}) \neq A_{\mathcal{U}}^{[i]}(\tilde{\mathcal{W}})$.
  - $A_{\mathcal{U}}^{[i]}(\mathcal{W})$ suffices to reconstruct the desired message.

Converse Proof: Main Body

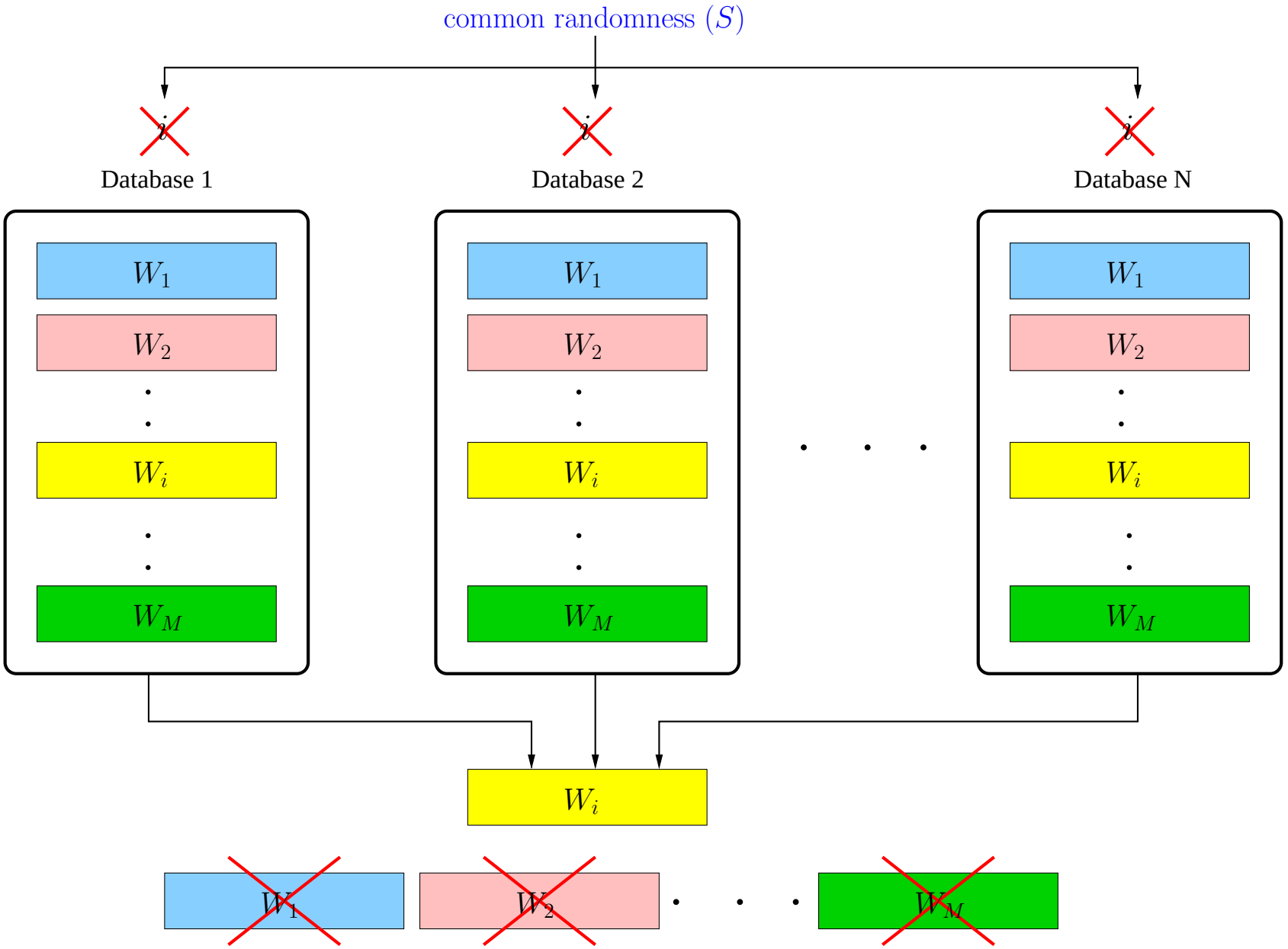
$$
\begin{aligned}
R &\leq \frac{L}{\sum_{n=1}^{N} H(A_n^{[i]}|\mathcal{Q})} \\
&= \underbrace{\frac{N-2B}{N}}_{\text{symmetry}} \cdot \underbrace{\frac{L}{\sum_{n\in\mathcal{U}} H(A_n^{[i]}(\mathcal{W})|\mathcal{Q})}}_{\text{responses from } \mathcal{U} \text{ suffice for decoding}} \\
&\leq \frac{N-2B}{N} \cdot \underbrace{C_T(N-2B)}_{\substack{\text{valid rate upper bounded by} \\ \text{the } T\text{-private capacity}}} \\
&= \frac{N-2B}{N} \cdot \underbrace{\frac{1-\frac{T}{N-2B}}{1-\left(\frac{T}{N-2B}\right)^M}}_{\substack{T\text{-private capacity} \\ \text{with } N-2B \text{ databases}}}
\end{aligned}
$$

Symmetric PIR (SPIR) [Sun-Jafar][6]



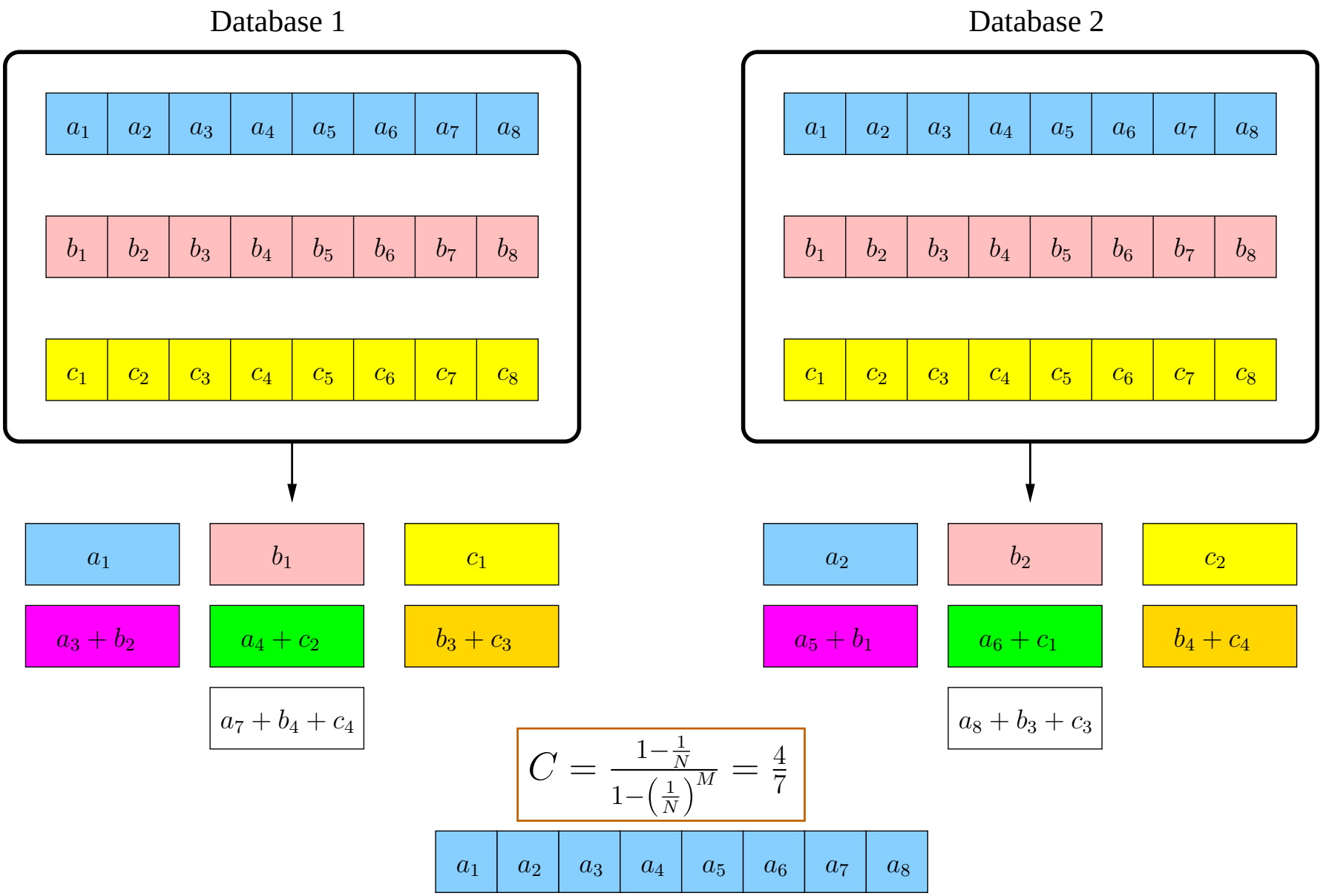
[6] H. Sun and S. Jafar. The capacity of symmetric private information retrieval. 2016. Available at arXiv:1606.08828.

Symmetric PIR (SPIR) [Sun-Jafar][6]

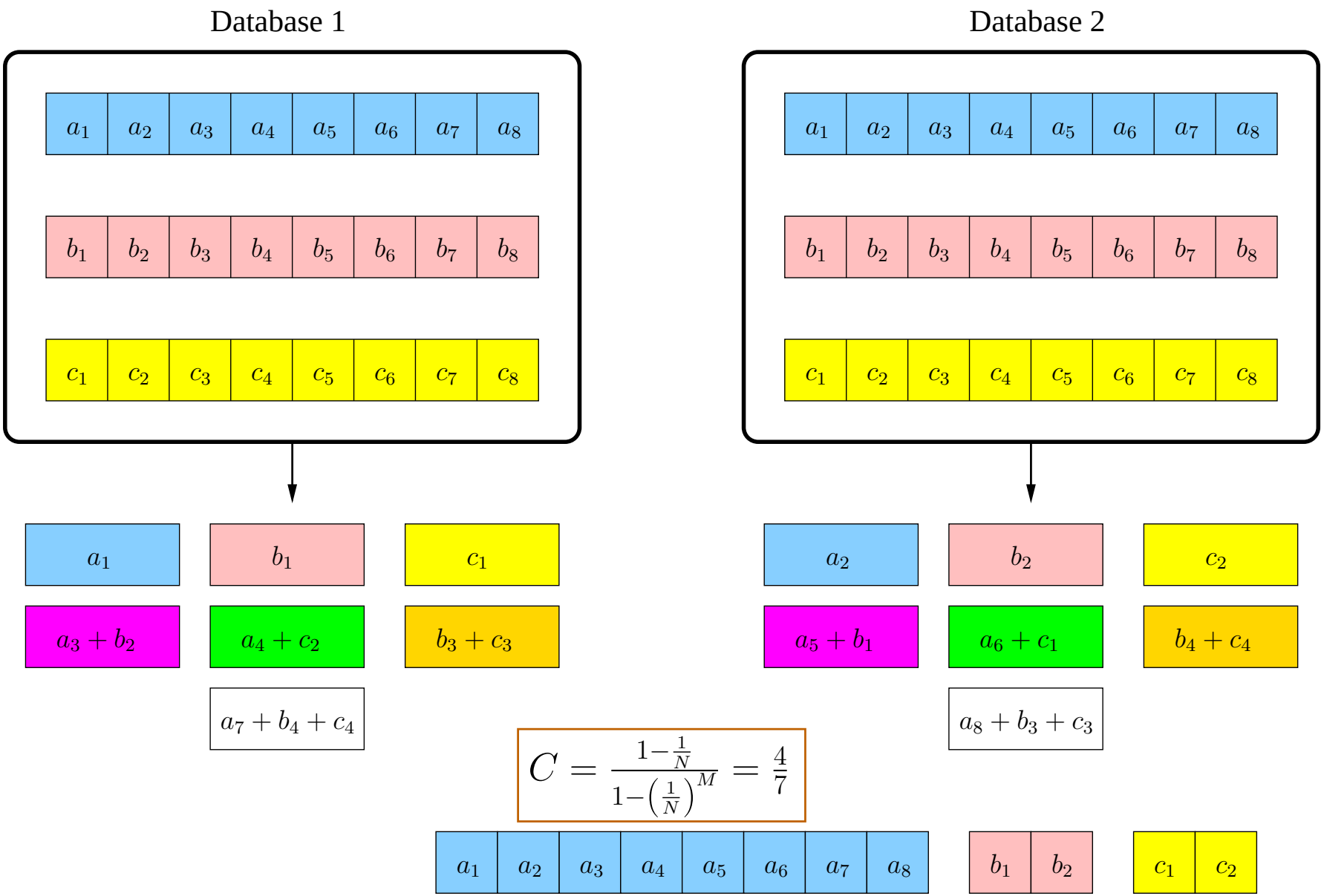


[6] H. Sun and S. Jafar. The capacity of symmetric private information retrieval. 2016. Available at arXiv:1606.08828.

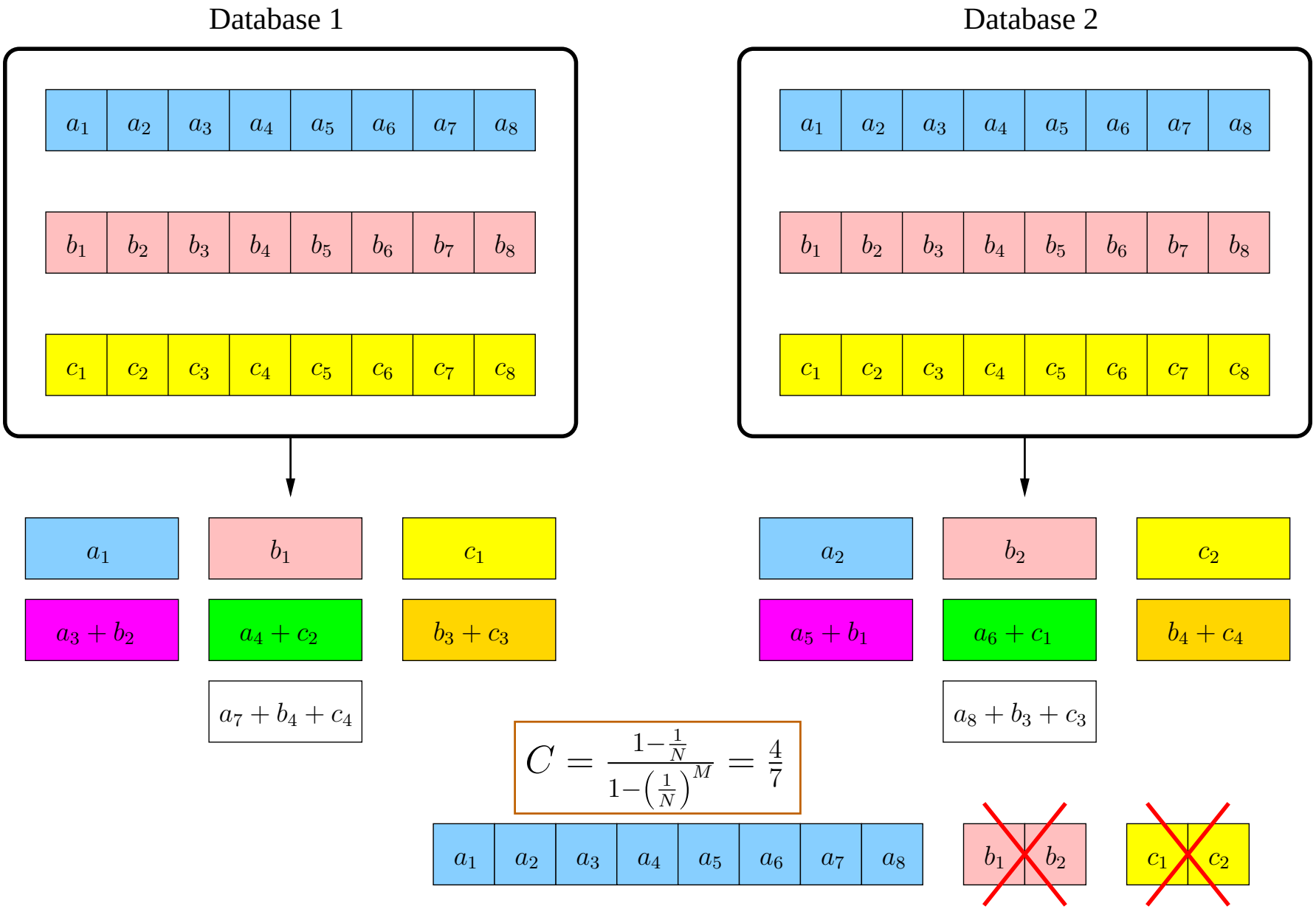
Example $M = 3, N = 2$: Private Retrieval Rate



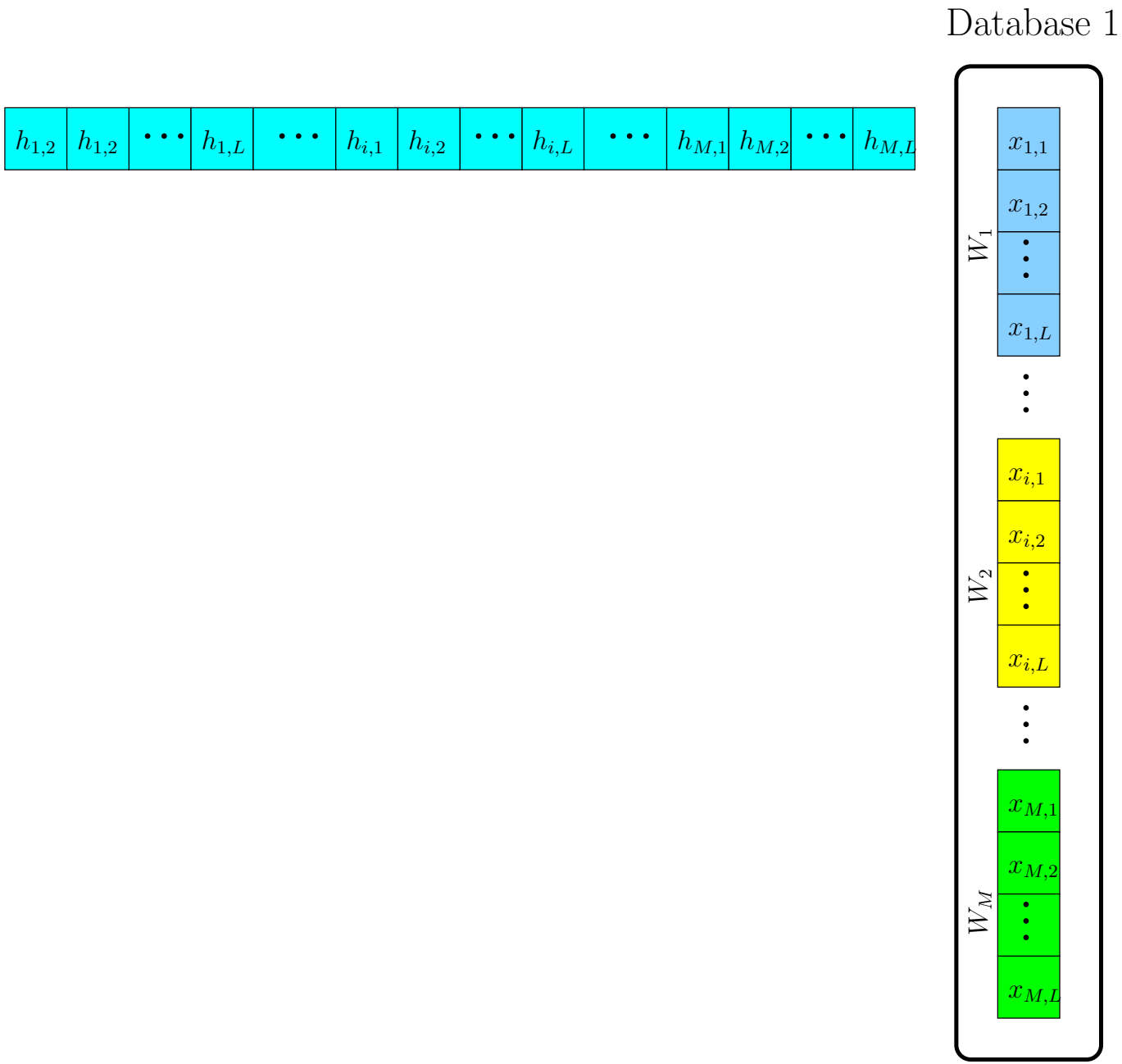
Example $M = 3, N = 2$: Private Retrieval Rate



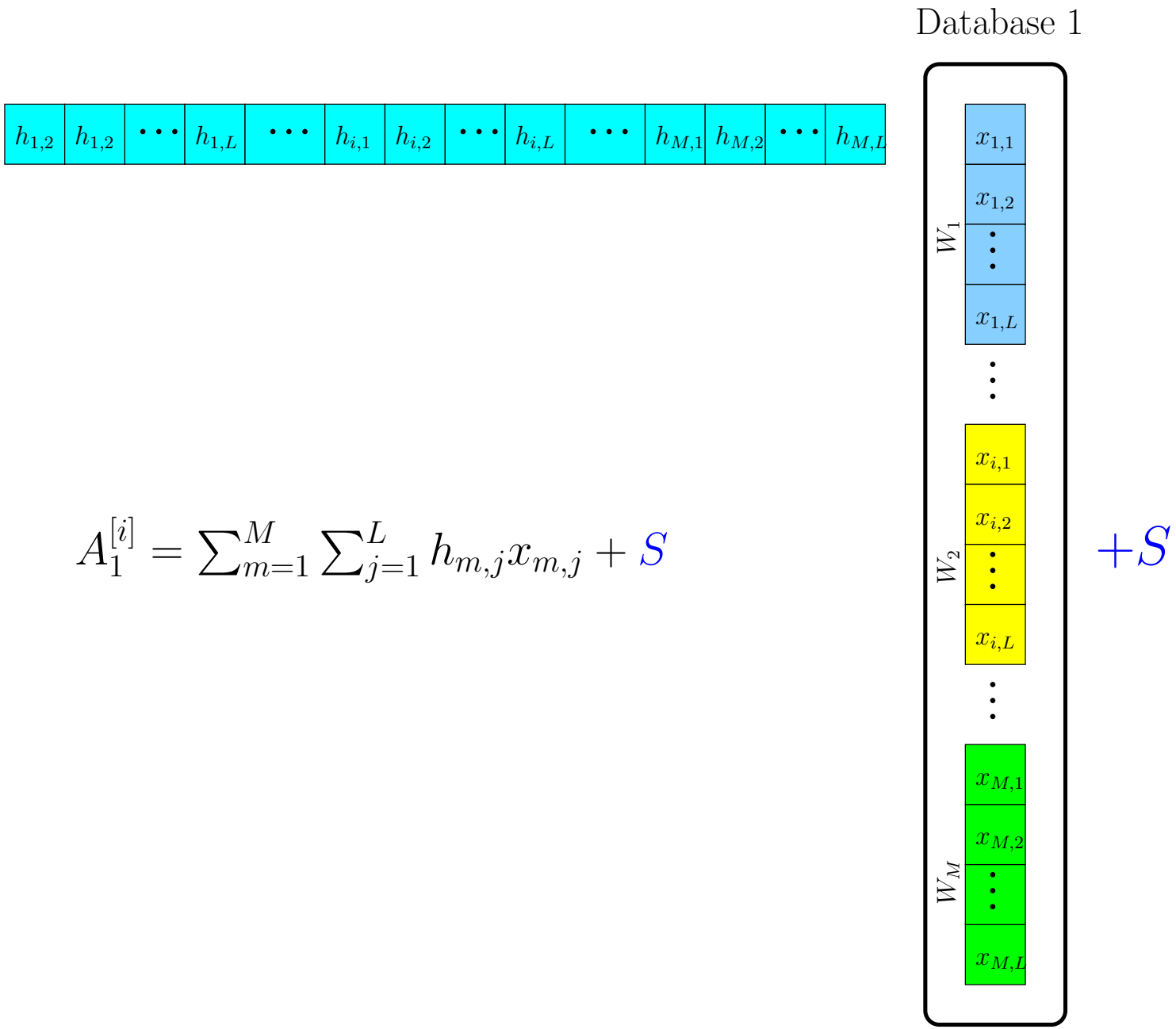
Example $M = 3, N = 2$: Private Retrieval Rate



Achievable Scheme

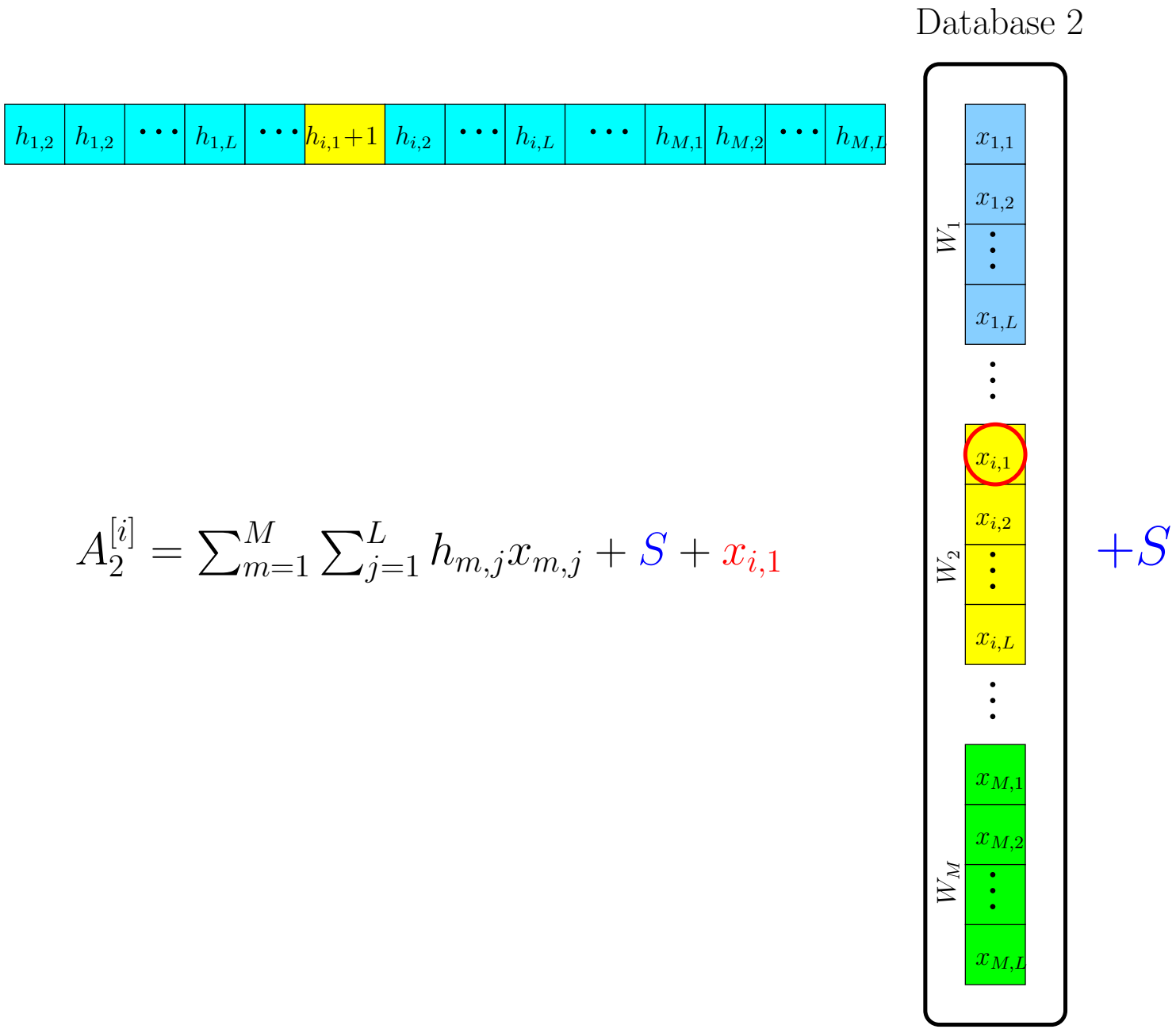


Achievable Scheme



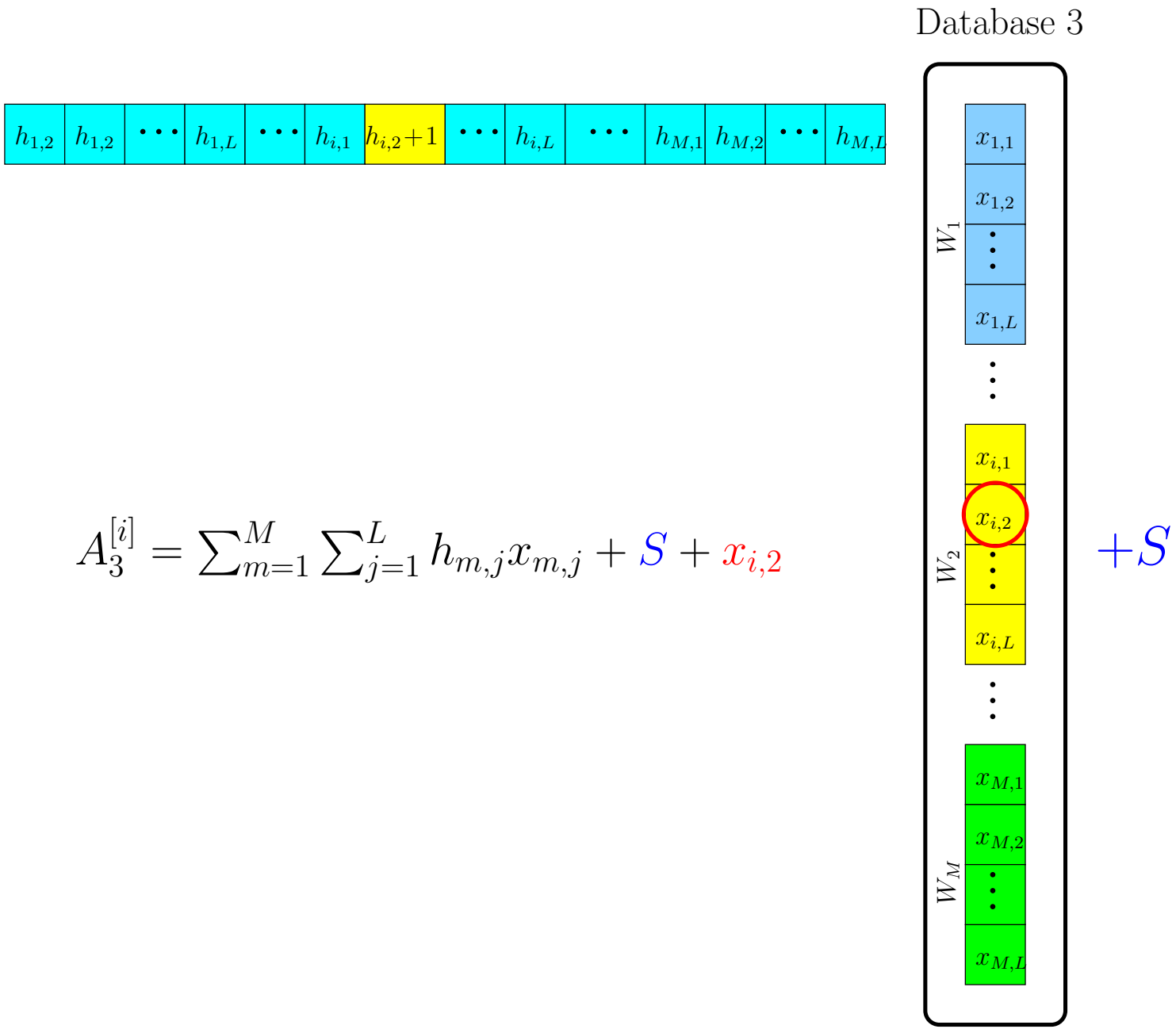
$$A_1^{[i]} = \sum_{m=1}^{M} \sum_{j=1}^{L} h_{m,j} x_{m,j} + S$$

Achievable Scheme



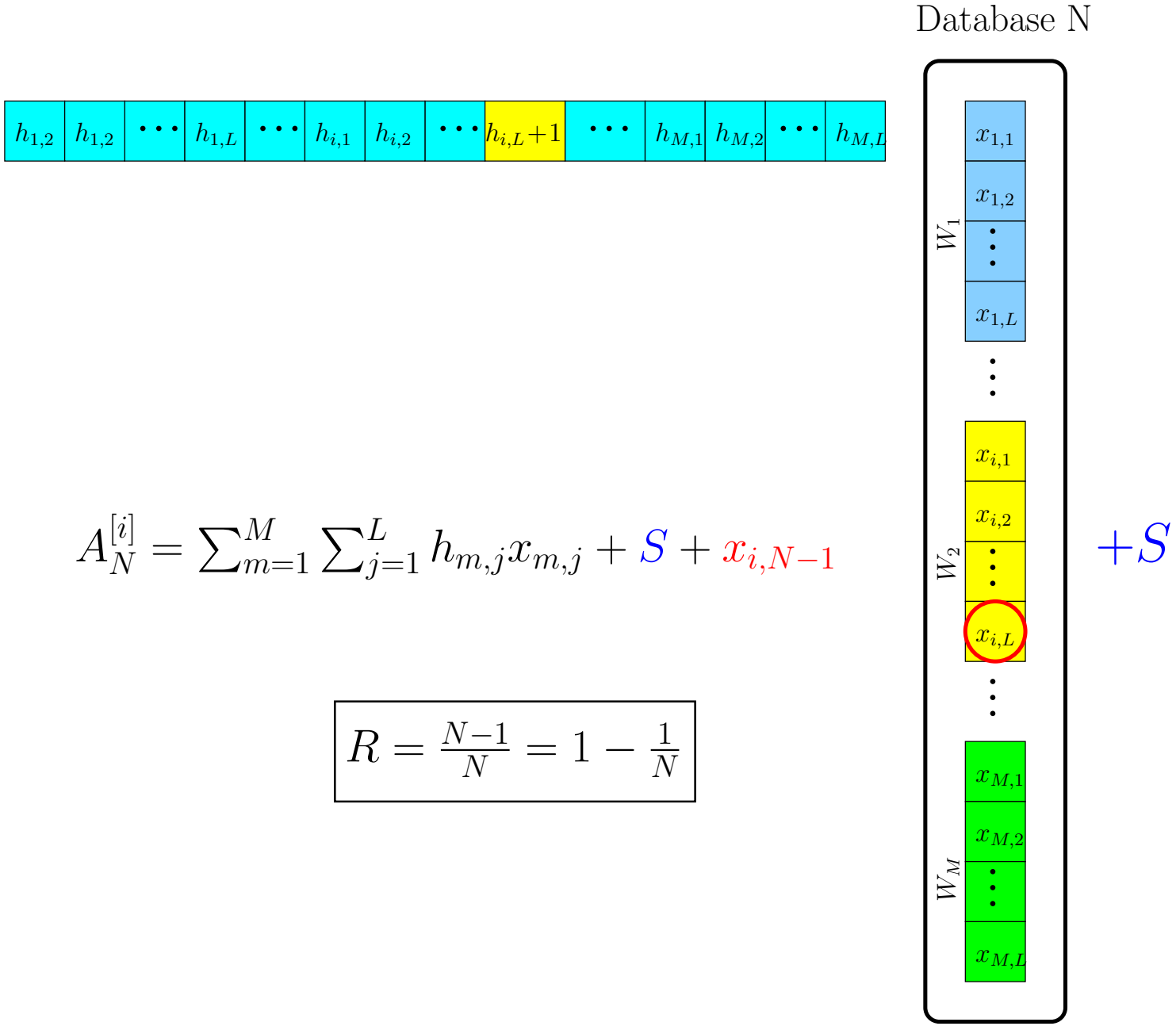
$$A_2^{[i]} = \sum_{m=1}^{M} \sum_{j=1}^{L} h_{m,j} x_{m,j} + S + x_{i,1}$$

Achievable Scheme



$$A_3^{[i]} = \sum_{m=1}^{M} \sum_{j=1}^{L} h_{m,j} x_{m,j} + S + x_{i,2}$$

Achievable Scheme



$$A_N^{[i]} = \sum_{m=1}^{M} \sum_{j=1}^{L} h_{m,j} x_{m,j} + S + x_{i,N-1}$$

$$\boxed{R = \frac{N-1}{N} = 1 - \frac{1}{N}}$$

Converse Proof: New Tools and Lemmas

- **New: Database privacy constraint**

$$I(W_1, \cdots, W_{i-1}, W_{i+1}, \cdots, W_M; A_{1:N}^{[i]}, Q_{1:N}^{[i]}, i) = I(W_{\bar{i}}; A_{1:N}^{[i]}, Q_{1:N}^{[i]}, i) = 0$$

- Secret key rate

$$\rho = \frac{H(S)}{L}$$

- **Lemma: Effect of conditioning**

$$H(A_n^{[m]} | W_m, Q_{1:N}^{[m]}) = H(A_n^{[m]} | Q_{1:N}^{[m]})$$

- **Intuition**
  - The uncertainty of the answer does not decrease after decoding W_m.
  - No induction for the SPIR model.
  - Consequence of the database privacy constraint.

Converse Proof: Main Body

- **Retrieval rate:** No need for induction

$$L = H(W_m) = \underbrace{H(W_m|Q_{1:N}^{[m]})}_{\text{independence}} - \underbrace{H(W_m|A_{1:N}^{[m]}, Q_{1:N}^{[m]})}_{=0 \text{ (reliability)}}$$

$$= I(W_m; A_{1:N}^{[m]}|Q_{1:N}^{[m]})$$

$$= H(A_{1:N}^{[m]}|Q_{1:N}^{[m]}) - H(A_{1:N}^{[m]}|W_m, Q_{1:N}^{[m]})$$

$$\leq H(A_{1:N}^{[m]}|Q_{1:N}^{[m]}) - H(A_n^{[m]}|W_m, Q_{1:N}^{[m]})$$

$$= H(A_{1:N}^{[m]}|Q_{1:N}^{[m]}) - H(A_n^{[m]}|Q_{1:N}^{[m]}) \quad \text{(effect of conditioning)}$$

- Adding over all n

$$NL \leq NH(A_{1:N}^{[m]}|Q_{1:N}^{[m]}) - \sum_{n=1}^{N} H(A_n^{[m]}|Q_{1:N}^{[m]})$$

$$\leq (N-1)H(A_{1:N}^{[m]}|Q_{1:N}^{[m]})$$

- Hence, the retrieval rate is upper bounded by

$$R \leq \frac{L}{\sum_{n=1}^{N} H(A_n^{[m]})} \leq \frac{L}{H(A_{1:N}^{[m]}|Q_{1:N}^{[m]})} = \frac{N-1}{N} = 1 - \frac{1}{N}$$

Converse Proof: Main Body (cont.)

- **Secret key rate:** Starting from the database privacy constraint

$$
\begin{aligned}
0 =& I(W_{\bar{m}}; A_{1:N}^{[m]}, Q_{1:N}^{[m]}) \\
=& I(W_{\bar{m}}; A_{1:N}^{[m]} | W_m, Q_{1:N}^{[m]}) \quad (W_{\bar{m}} \text{ are independent of } W_m, Q_{1:N}^{[m]}) \\
\geq& I(W_{\bar{m}}; A_n^{[m]} | W_m, Q_{1:N}^{[m]}) \\
=& H(A_n^{[m]} | W_m, Q_{1:N}^{[m]}) - \underbrace{H(A_n^{[m]} | W_{1:M}, Q_{1:N}^{[m]})}_{\leq H(S)} \\
\geq& H(A_n^{[m]} | Q_{1:N}^{[m]}) - H(S) \quad (\text{effect of conditioning})
\end{aligned}
$$

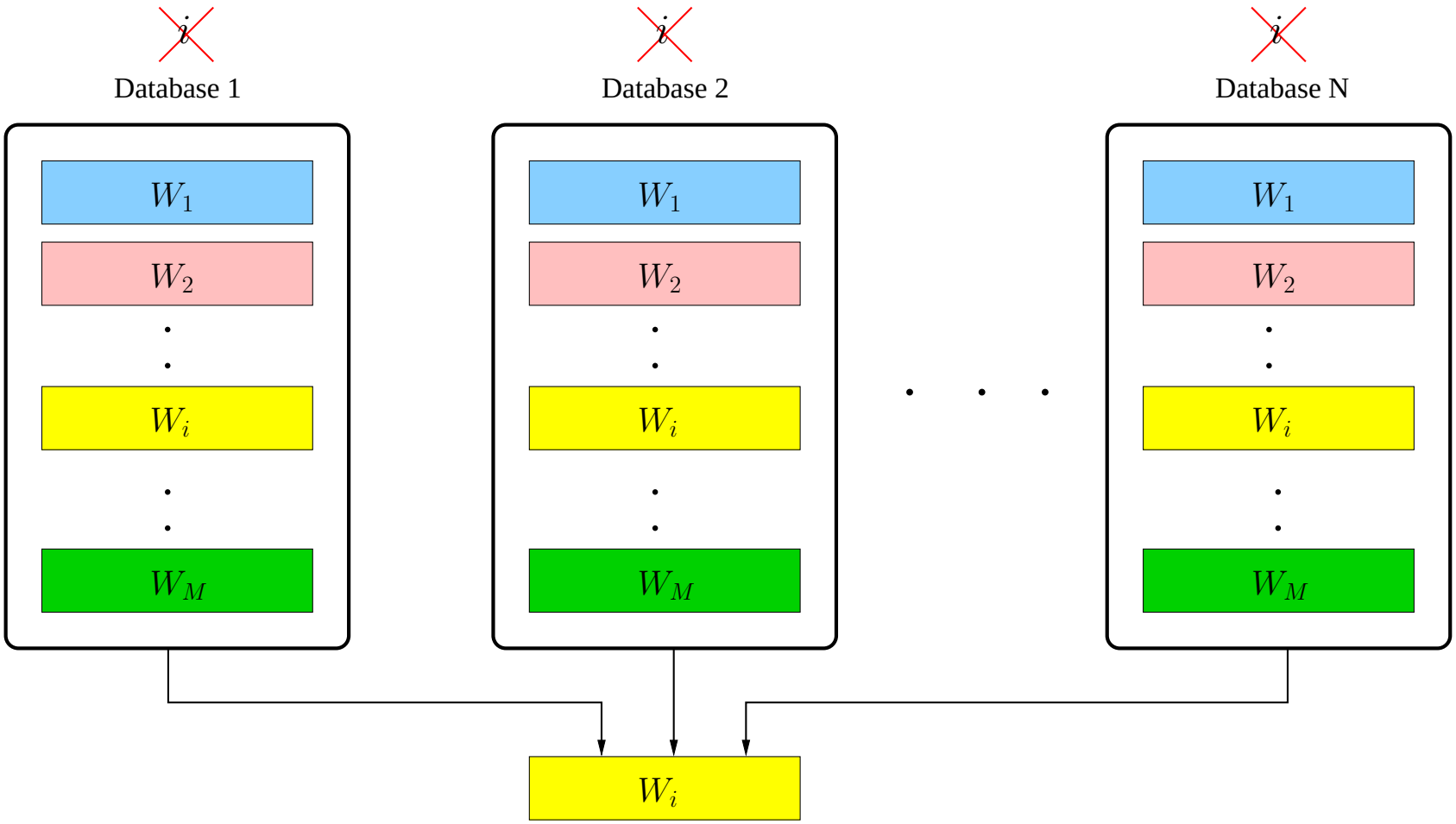
- Hence, $H(S) \geq H(A_n^{[m]} | Q_{1:N}^{[m]})$.
- Adding over all n

$$
\begin{aligned}
NH(S) &\geq \sum_{n=1}^{N} H(A_n^{[m]} | Q_{1:N}^{[m]}) \\
&\geq H(A_{1:N}^{[m]} | Q_{1:N}^{[m]}) \\
&\geq \frac{NL}{N-1}
\end{aligned}
$$

- Secret key rate is lower bounded by

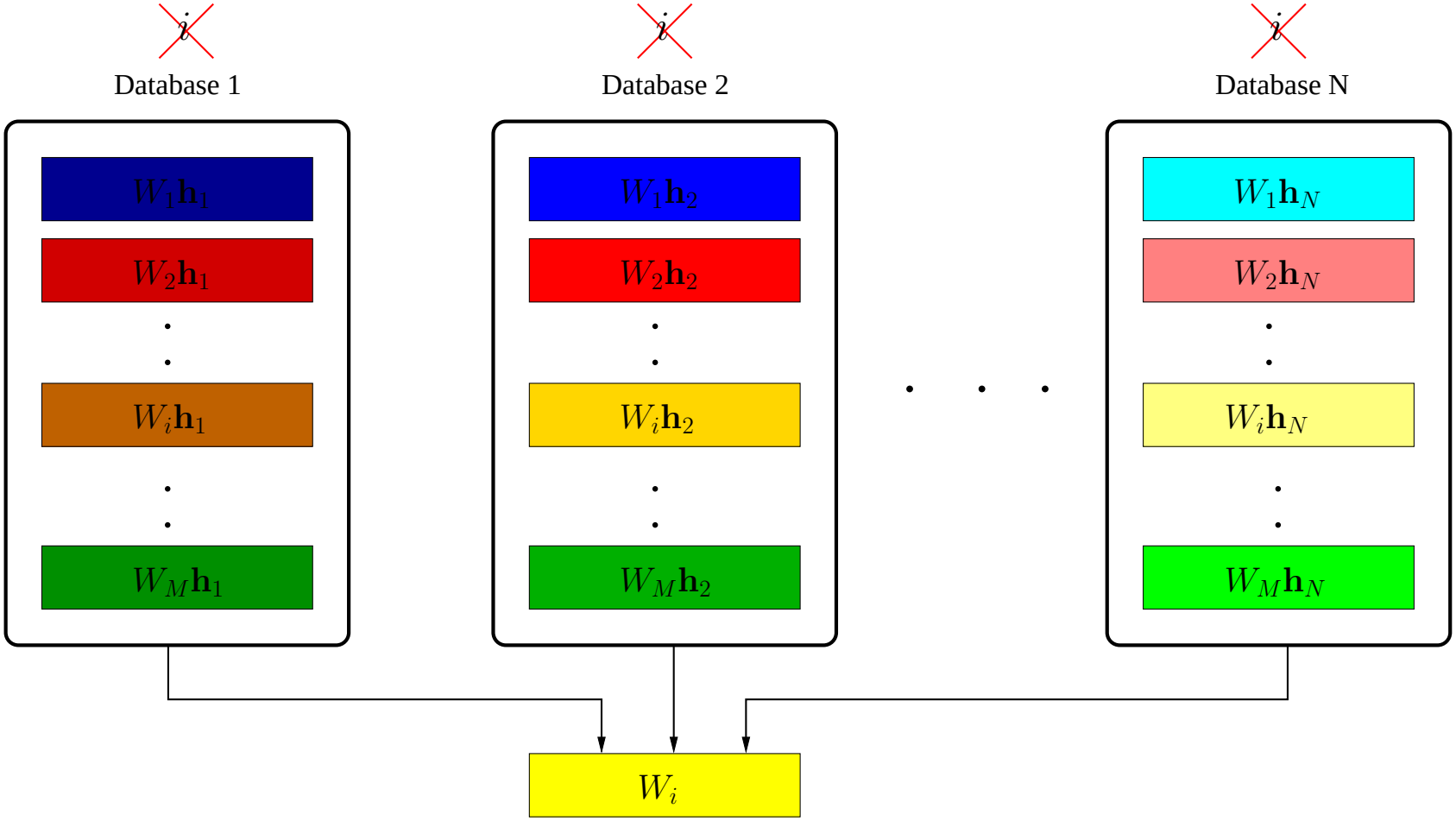
$$
\rho = \frac{H(S)}{L} \geq \frac{1}{N-1}
$$

PIR with Coded Databases (CPIR) [Banawan-Ulukus][7]



[7] K. Banawan and S. Ulukus, The Capacity of Private Information Retrieval from Coded Databases, IEEE Trans. on Information Theory, 64(3):1945-1956, March 2018.

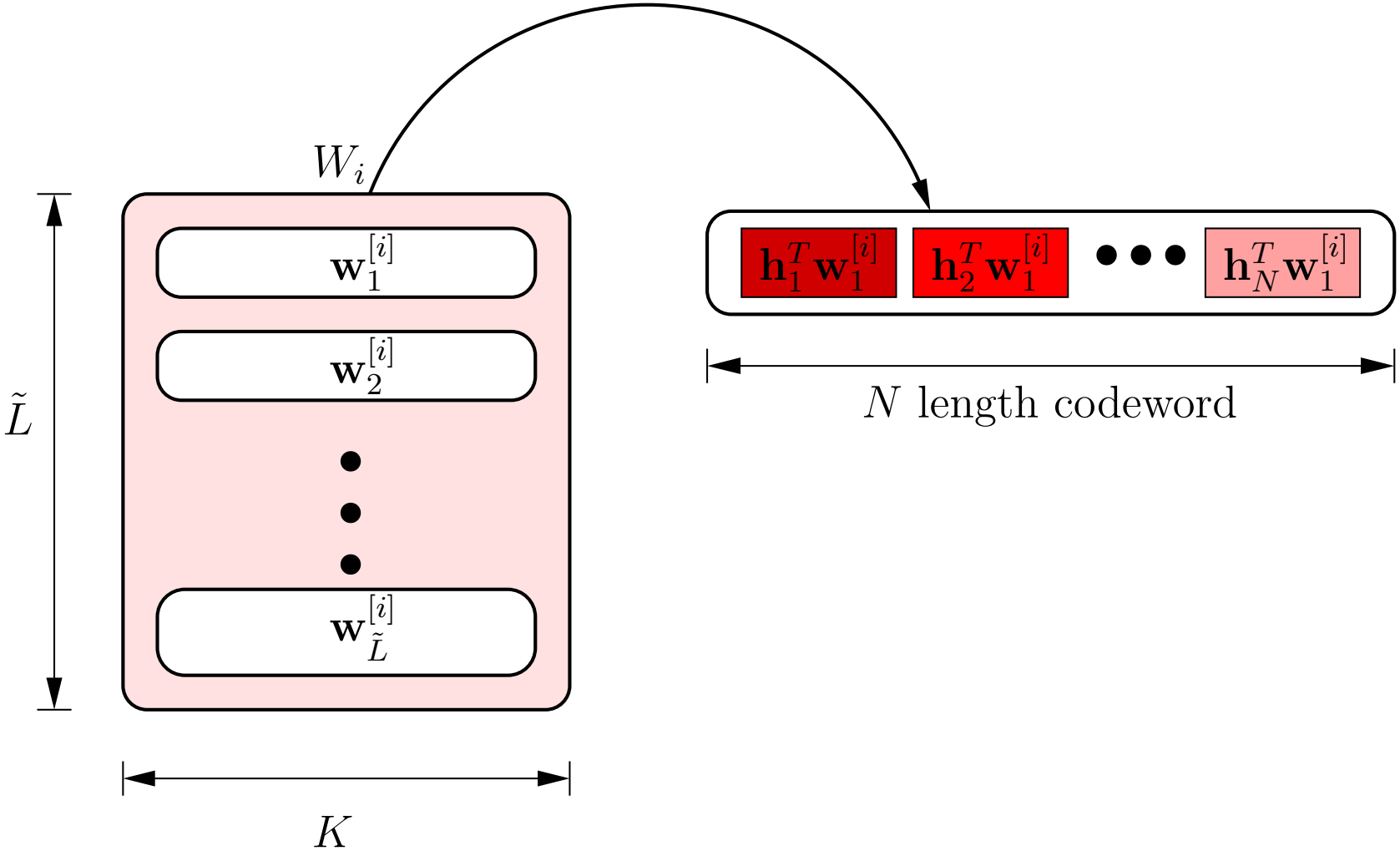
PIR with Coded Databases (CPIR) [Banawan-Ulukus][7]



[7] K. Banawan and S. Ulukus, The Capacity of Private Information Retrieval from Coded Databases, IEEE Trans. on Information Theory, 64(3):1945-1956, March 2018.

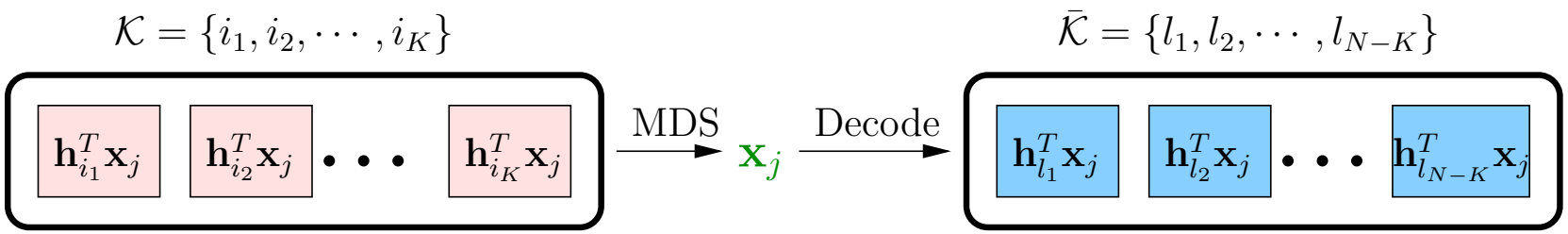
Code Structure

coding every row via generator matrix $\mathbf{H} \in \mathbb{F}_q^{K \times N}$



Achievable Scheme: New Ingredients

- MDS property: Any K elements of a row decode the entire row.



- Scheme is performed in K repetitions.

- **For undesired symbols:**
  - Download K symbols from same row from K different databases.
  - Use this row in the other $(N - K)$ databases as side information.

- **For desired symbols:**
  - The start of the scheme shifts circularly each repetition.
  - Symbols are decoded after K repetitions.

CPIR Example: $M = 2$, $(3, 2)$ code

		DB1	DB2	DB3
repetition 1	round 1	$\mathbf{h}_1^T \mathbf{x}_1^{[1]}$ $\mathbf{h}_1^T \mathbf{x}_2^{[1]}$ $\mathbf{h}_1^T \mathbf{x}_1^{[2]}$ $\mathbf{h}_1^T \mathbf{x}_2^{[2]}$	$\mathbf{h}_2^T \mathbf{x}_3^{[1]}$ $\mathbf{h}_2^T \mathbf{x}_4^{[1]}$ $\mathbf{h}_2^T \mathbf{x}_1^{[2]}$ $\mathbf{h}_2^T \mathbf{x}_3^{[2]}$	$\mathbf{h}_3^T \mathbf{x}_5^{[1]}$ $\mathbf{h}_3^T \mathbf{x}_6^{[1]}$ $\mathbf{h}_3^T \mathbf{x}_2^{[2]}$ $\mathbf{h}_3^T \mathbf{x}_3^{[2]}$
	rd.2	$\mathbf{h}_1^T (\mathbf{x}_7^{[1]} + \mathbf{x}_3^{[2]})$	$\mathbf{h}_2^T (\mathbf{x}_8^{[1]} + \mathbf{x}_2^{[2]})$	$\mathbf{h}_3^T (\mathbf{x}_9^{[1]} + \mathbf{x}_1^{[2]})$
repetition 2	round 1	$\mathbf{h}_1^T \mathbf{x}_5^{[1]}$ $\mathbf{h}_1^T \mathbf{x}_6^{[1]}$ $\mathbf{h}_1^T \mathbf{x}_4^{[2]}$ $\mathbf{h}_1^T \mathbf{x}_5^{[2]}$	$\mathbf{h}_2^T \mathbf{x}_1^{[1]}$ $\mathbf{h}_2^T \mathbf{x}_2^{[1]}$ $\mathbf{h}_2^T \mathbf{x}_4^{[2]}$ $\mathbf{h}_2^T \mathbf{x}_6^{[2]}$	$\mathbf{h}_3^T \mathbf{x}_3^{[1]}$ $\mathbf{h}_3^T \mathbf{x}_4^{[1]}$ $\mathbf{h}_3^T \mathbf{x}_5^{[2]}$ $\mathbf{h}_3^T \mathbf{x}_6^{[2]}$
	rd.2	$\mathbf{h}_1^T (\mathbf{x}_9^{[1]} + \mathbf{x}_6^{[2]})$	$\mathbf{h}_2^T (\mathbf{x}_7^{[1]} + \mathbf{x}_5^{[2]})$	$\mathbf{h}_3^T (\mathbf{x}_8^{[1]} + \mathbf{x}_4^{[2]})$

CPIR Example: $M = 2$, $(3, 2)$ code

		DB1	DB2	DB3
repetition 1	round 1	$\mathbf{h}_1^T \mathbf{x}_1^{[1]}$ $\mathbf{h}_1^T \mathbf{x}_2^{[1]}$ $\mathbf{h}_1^T \mathbf{x}_1^{[2]}$ $\mathbf{h}_1^T \mathbf{x}_2^{[2]}$	$\mathbf{h}_2^T \mathbf{x}_3^{[1]}$ $\mathbf{h}_2^T \mathbf{x}_4^{[1]}$ $\mathbf{h}_2^T \mathbf{x}_1^{[2]}$ $\mathbf{h}_2^T \mathbf{x}_3^{[2]}$	$\mathbf{h}_3^T \mathbf{x}_5^{[1]}$ $\mathbf{h}_3^T \mathbf{x}_6^{[1]}$ $\mathbf{h}_3^T \mathbf{x}_2^{[2]}$ $\mathbf{h}_3^T \mathbf{x}_3^{[2]}$
	rd.2	$\mathbf{h}_1^T (\mathbf{x}_7^{[1]} + \mathbf{x}_3^{[2]})$	$\mathbf{h}_2^T (\mathbf{x}_8^{[1]} + \mathbf{x}_2^{[2]})$	$\mathbf{h}_3^T (\mathbf{x}_9^{[1]} + \mathbf{x}_1^{[2]})$
repetition 2	round 1	$\mathbf{h}_1^T \mathbf{x}_5^{[1]}$ $\mathbf{h}_1^T \mathbf{x}_6^{[1]}$ $\mathbf{h}_1^T \mathbf{x}_4^{[2]}$ $\mathbf{h}_1^T \mathbf{x}_5^{[2]}$	$\mathbf{h}_2^T \mathbf{x}_1^{[1]}$ $\mathbf{h}_2^T \mathbf{x}_2^{[1]}$ $\mathbf{h}_2^T \mathbf{x}_4^{[2]}$ $\mathbf{h}_2^T \mathbf{x}_6^{[2]}$	$\mathbf{h}_3^T \mathbf{x}_3^{[1]}$ $\mathbf{h}_3^T \mathbf{x}_4^{[1]}$ $\mathbf{h}_3^T \mathbf{x}_5^{[2]}$ $\mathbf{h}_3^T \mathbf{x}_6^{[2]}$
	rd.2	$\mathbf{h}_1^T (\mathbf{x}_9^{[1]} + \mathbf{x}_6^{[2]})$	$\mathbf{h}_2^T (\mathbf{x}_7^{[1]} + \mathbf{x}_5^{[2]})$	$\mathbf{h}_3^T (\mathbf{x}_8^{[1]} + \mathbf{x}_4^{[2]})$

CPIR Example: $M = 2$, $(3, 2)$ code

		DB1	DB2	DB3
repetition 1	round 1	$\mathbf{h}_1^T \mathbf{x}_1^{[1]}$ $\mathbf{h}_1^T \mathbf{x}_2^{[1]}$ $\mathbf{h}_1^T \mathbf{x}_1^{[2]}$ $\mathbf{h}_1^T \mathbf{x}_2^{[2]}$	$\mathbf{h}_2^T \mathbf{x}_3^{[1]}$ $\mathbf{h}_2^T \mathbf{x}_4^{[1]}$ $\mathbf{h}_2^T \mathbf{x}_1^{[2]}$ $\mathbf{h}_2^T \mathbf{x}_3^{[2]}$	$\mathbf{h}_3^T \mathbf{x}_5^{[1]}$ $\mathbf{h}_3^T \mathbf{x}_6^{[1]}$ $\mathbf{h}_3^T \mathbf{x}_2^{[2]}$ $\mathbf{h}_3^T \mathbf{x}_3^{[2]}$
	rd.2	$\mathbf{h}_1^T (\mathbf{x}_7^{[1]} + \mathbf{x}_3^{[2]})$	$\mathbf{h}_2^T (\mathbf{x}_8^{[1]} + \mathbf{x}_2^{[2]})$	$\mathbf{h}_3^T (\mathbf{x}_9^{[1]} + \mathbf{x}_1^{[2]})$
repetition 2	round 1	$\mathbf{h}_1^T \mathbf{x}_5^{[1]}$ $\mathbf{h}_1^T \mathbf{x}_6^{[1]}$ $\mathbf{h}_1^T \mathbf{x}_4^{[2]}$ $\mathbf{h}_1^T \mathbf{x}_5^{[2]}$	$\mathbf{h}_2^T \mathbf{x}_1^{[1]}$ $\mathbf{h}_2^T \mathbf{x}_2^{[1]}$ $\mathbf{h}_2^T \mathbf{x}_4^{[2]}$ $\mathbf{h}_2^T \mathbf{x}_6^{[2]}$	$\mathbf{h}_3^T \mathbf{x}_3^{[1]}$ $\mathbf{h}_3^T \mathbf{x}_4^{[1]}$ $\mathbf{h}_3^T \mathbf{x}_5^{[2]}$ $\mathbf{h}_3^T \mathbf{x}_6^{[2]}$
	rd.2	$\mathbf{h}_1^T (\mathbf{x}_9^{[1]} + \mathbf{x}_6^{[2]})$	$\mathbf{h}_2^T (\mathbf{x}_7^{[1]} + \mathbf{x}_5^{[2]})$	$\mathbf{h}_3^T (\mathbf{x}_8^{[1]} + \mathbf{x}_4^{[2]})$

CPIR Example: $M = 2$, $(3, 2)$ code

		DB1	DB2	DB3
repetition 1	round 1	$\mathbf{h}_1^T \mathbf{x}_1^{[1]}$ $\mathbf{h}_1^T \mathbf{x}_2^{[1]}$ $\mathbf{h}_1^T \mathbf{x}_1^{[2]}$ $\mathbf{h}_1^T \mathbf{x}_2^{[2]}$	$\mathbf{h}_2^T \mathbf{x}_3^{[1]}$ $\mathbf{h}_2^T \mathbf{x}_4^{[1]}$ $\mathbf{h}_2^T \mathbf{x}_1^{[2]}$ $\mathbf{h}_2^T \mathbf{x}_3^{[2]}$	$\mathbf{h}_3^T \mathbf{x}_5^{[1]}$ $\mathbf{h}_3^T \mathbf{x}_6^{[1]}$ $\mathbf{h}_3^T \mathbf{x}_2^{[2]}$ $\mathbf{h}_3^T \mathbf{x}_3^{[2]}$
	rd.2	$\mathbf{h}_1^T (\mathbf{x}_7^{[1]} + \mathbf{x}_3^{[2]})$	$\mathbf{h}_2^T (\mathbf{x}_8^{[1]} + \mathbf{x}_2^{[2]})$	$\mathbf{h}_3^T (\mathbf{x}_9^{[1]} + \mathbf{x}_1^{[2]})$
repetition 2	round 1	$\mathbf{h}_1^T \mathbf{x}_5^{[1]}$ $\mathbf{h}_1^T \mathbf{x}_6^{[1]}$ $\mathbf{h}_1^T \mathbf{x}_4^{[2]}$ $\mathbf{h}_1^T \mathbf{x}_5^{[2]}$	$\mathbf{h}_2^T \mathbf{x}_1^{[1]}$ $\mathbf{h}_2^T \mathbf{x}_2^{[1]}$ $\mathbf{h}_2^T \mathbf{x}_4^{[2]}$ $\mathbf{h}_2^T \mathbf{x}_6^{[2]}$	$\mathbf{h}_3^T \mathbf{x}_3^{[1]}$ $\mathbf{h}_3^T \mathbf{x}_4^{[1]}$ $\mathbf{h}_3^T \mathbf{x}_5^{[2]}$ $\mathbf{h}_3^T \mathbf{x}_6^{[2]}$
	rd.2	$\mathbf{h}_1^T (\mathbf{x}_9^{[1]} + \mathbf{x}_6^{[2]})$	$\mathbf{h}_2^T (\mathbf{x}_7^{[1]} + \mathbf{x}_5^{[2]})$	$\mathbf{h}_3^T (\mathbf{x}_8^{[1]} + \mathbf{x}_4^{[2]})$

CPIR Example: $M = 2$, $(3, 2)$ code

		DB1	DB2	DB3
repetition 1	round 1	$\mathbf{h}_1^T \mathbf{x}_1^{[1]}$ $\mathbf{h}_1^T \mathbf{x}_2^{[1]}$ $\mathbf{h}_1^T \mathbf{x}_1^{[2]}$ $\mathbf{h}_1^T \mathbf{x}_2^{[2]}$	$\mathbf{h}_2^T \mathbf{x}_3^{[1]}$ $\mathbf{h}_2^T \mathbf{x}_4^{[1]}$ $\mathbf{h}_2^T \mathbf{x}_1^{[2]}$ $\mathbf{h}_2^T \mathbf{x}_3^{[2]}$	$\mathbf{h}_3^T \mathbf{x}_5^{[1]}$ $\mathbf{h}_3^T \mathbf{x}_6^{[1]}$ $\mathbf{h}_3^T \mathbf{x}_2^{[2]}$ $\mathbf{h}_3^T \mathbf{x}_3^{[2]}$
	rd.2	$\mathbf{h}_1^T (\mathbf{x}_7^{[1]} + \mathbf{x}_3^{[2]})$	$\mathbf{h}_2^T (\mathbf{x}_8^{[1]} + \mathbf{x}_2^{[2]})$	$\mathbf{h}_3^T (\mathbf{x}_9^{[1]} + \mathbf{x}_1^{[2]})$
repetition 2	round 1	$\mathbf{h}_1^T \mathbf{x}_5^{[1]}$ $\mathbf{h}_1^T \mathbf{x}_6^{[1]}$ $\mathbf{h}_1^T \mathbf{x}_4^{[2]}$ $\mathbf{h}_1^T \mathbf{x}_5^{[2]}$	$\mathbf{h}_2^T \mathbf{x}_1^{[1]}$ $\mathbf{h}_2^T \mathbf{x}_2^{[1]}$ $\mathbf{h}_2^T \mathbf{x}_4^{[2]}$ $\mathbf{h}_2^T \mathbf{x}_6^{[2]}$	$\mathbf{h}_3^T \mathbf{x}_3^{[1]}$ $\mathbf{h}_3^T \mathbf{x}_4^{[1]}$ $\mathbf{h}_3^T \mathbf{x}_5^{[2]}$ $\mathbf{h}_3^T \mathbf{x}_6^{[2]}$
	rd.2	$\mathbf{h}_1^T (\mathbf{x}_9^{[1]} + \mathbf{x}_6^{[2]})$	$\mathbf{h}_2^T (\mathbf{x}_7^{[1]} + \mathbf{x}_5^{[2]})$	$\mathbf{h}_3^T (\mathbf{x}_8^{[1]} + \mathbf{x}_4^{[2]})$

CPIR Example: $M = 2$, $(3, 2)$ code

		DB1	DB2	DB3
repetition 1	round 1	$\mathbf{h}_1^T \mathbf{x}_1^{[1]}$ $\mathbf{h}_1^T \mathbf{x}_2^{[1]}$ $\mathbf{h}_1^T \mathbf{x}_1^{[2]}$ $\mathbf{h}_1^T \mathbf{x}_2^{[2]}$	$\mathbf{h}_2^T \mathbf{x}_3^{[1]}$ $\mathbf{h}_2^T \mathbf{x}_4^{[1]}$ $\mathbf{h}_2^T \mathbf{x}_1^{[2]}$ $\mathbf{h}_2^T \mathbf{x}_3^{[2]}$	$\mathbf{h}_3^T \mathbf{x}_5^{[1]}$ $\mathbf{h}_3^T \mathbf{x}_6^{[1]}$ $\mathbf{h}_3^T \mathbf{x}_2^{[2]}$ $\mathbf{h}_3^T \mathbf{x}_3^{[2]}$
	rd.2	$\mathbf{h}_1^T (\mathbf{x}_7^{[1]} + \cancel{\mathbf{x}_3^{[2]}})$	$\mathbf{h}_2^T (\mathbf{x}_8^{[1]} + \cancel{\mathbf{x}_2^{[2]}})$	$\mathbf{h}_3^T (\mathbf{x}_9^{[1]} + \cancel{\mathbf{x}_1^{[2]}})$
repetition 2	round 1	$\mathbf{h}_1^T \mathbf{x}_5^{[1]}$ $\mathbf{h}_1^T \mathbf{x}_6^{[1]}$ $\mathbf{h}_1^T \mathbf{x}_4^{[2]}$ $\mathbf{h}_1^T \mathbf{x}_5^{[2]}$	$\mathbf{h}_2^T \mathbf{x}_1^{[1]}$ $\mathbf{h}_2^T \mathbf{x}_2^{[1]}$ $\mathbf{h}_2^T \mathbf{x}_4^{[2]}$ $\mathbf{h}_2^T \mathbf{x}_6^{[2]}$	$\mathbf{h}_3^T \mathbf{x}_3^{[1]}$ $\mathbf{h}_3^T \mathbf{x}_4^{[1]}$ $\mathbf{h}_3^T \mathbf{x}_5^{[2]}$ $\mathbf{h}_3^T \mathbf{x}_6^{[2]}$
	rd.2	$\mathbf{h}_1^T (\mathbf{x}_9^{[1]} + \cancel{\mathbf{x}_6^{[2]}})$	$\mathbf{h}_2^T (\mathbf{x}_7^{[1]} + \cancel{\mathbf{x}_5^{[2]}})$	$\mathbf{h}_3^T (\mathbf{x}_8^{[1]} + \cancel{\mathbf{x}_4^{[2]}})$

CPIR Example: $M = 2$, $(3, 2)$ code

		DB1	DB2	DB3
repetition 1	round 1	$\mathbf{h}_1^T \mathbf{x}_1^{[1]}$ $\mathbf{h}_1^T \mathbf{x}_2^{[1]}$ $\mathbf{h}_1^T \mathbf{x}_1^{[2]}$ $\mathbf{h}_1^T \mathbf{x}_2^{[2]}$	$\mathbf{h}_2^T \mathbf{x}_3^{[1]}$ $\mathbf{h}_2^T \mathbf{x}_4^{[1]}$ $\mathbf{h}_2^T \mathbf{x}_1^{[2]}$ $\mathbf{h}_2^T \mathbf{x}_3^{[2]}$	$\mathbf{h}_3^T \mathbf{x}_5^{[1]}$ $\mathbf{h}_3^T \mathbf{x}_6^{[1]}$ $\mathbf{h}_3^T \mathbf{x}_2^{[2]}$ $\mathbf{h}_3^T \mathbf{x}_3^{[2]}$
	rd.2	$\mathbf{h}_1^T \mathbf{x}_7^{[1]}$	$\mathbf{h}_2^T \mathbf{x}_8^{[1]}$	$\mathbf{h}_3^T \mathbf{x}_9^{[1]}$
repetition 2	round 1	$\mathbf{h}_1^T \mathbf{x}_5^{[1]}$ $\mathbf{h}_1^T \mathbf{x}_6^{[1]}$ $\mathbf{h}_1^T \mathbf{x}_4^{[2]}$ $\mathbf{h}_1^T \mathbf{x}_5^{[2]}$	$\mathbf{h}_2^T \mathbf{x}_1^{[1]}$ $\mathbf{h}_2^T \mathbf{x}_2^{[1]}$ $\mathbf{h}_2^T \mathbf{x}_4^{[2]}$ $\mathbf{h}_2^T \mathbf{x}_6^{[2]}$	$\mathbf{h}_3^T \mathbf{x}_3^{[1]}$ $\mathbf{h}_3^T \mathbf{x}_4^{[1]}$ $\mathbf{h}_3^T \mathbf{x}_5^{[2]}$ $\mathbf{h}_3^T \mathbf{x}_6^{[2]}$
	rd.2	$\mathbf{h}_1^T \mathbf{x}_9^{[1]}$	$\mathbf{h}_2^T \mathbf{x}_7^{[1]}$	$\mathbf{h}_3^T \mathbf{x}_8^{[1]}$

$$R = \frac{1 - \frac{K}{N}}{1 - (\frac{K}{N})^M} = \frac{1 - \frac{2}{3}}{1 - (\frac{2}{3})^2} = \frac{9 * 2}{10 * 3} = \frac{3}{5}$$

Revisiting Capacity Formulas

- Classical PIR:

$$C_{\text{PIR}} = \frac{1 - \frac{1}{N}}{1 - (\frac{1}{N})^M}$$

- T-colluding PIR:

$$C_{\text{COL}} = \frac{1 - \frac{T}{N}}{1 - (\frac{T}{N})^M}$$

- U-robust PIR:

$$C_{\text{ROB}} = \frac{1 - \frac{T}{N-U}}{1 - (\frac{T}{N-U})^M}$$

- B-Byzantine PIR:

$$C_{\text{BYZ}} = \frac{N-2B}{N} \cdot \frac{1 - \frac{T}{N-2B}}{1 - (\frac{T}{N-2B})^M}$$

- Symmetric PIR:

$$C_{\text{SYM}} = 1 - \frac{1}{N}$$

- (N, K) MDS-coded PIR:

$$C_{\text{MDS}} = \frac{1 - \frac{K}{N}}{1 - (\frac{K}{N})^M}$$

Converse Proof: New Tools and Lemmas

- For MDS-code, contents of any K databases are linearly independent.

- **Lemma: Independence of any K answers**
  For any $\mathcal{K} \subset \{1, \cdots, N\}$ such that $|\mathcal{K}| = K$, and subset $W_{\mathcal{S}}$ messages

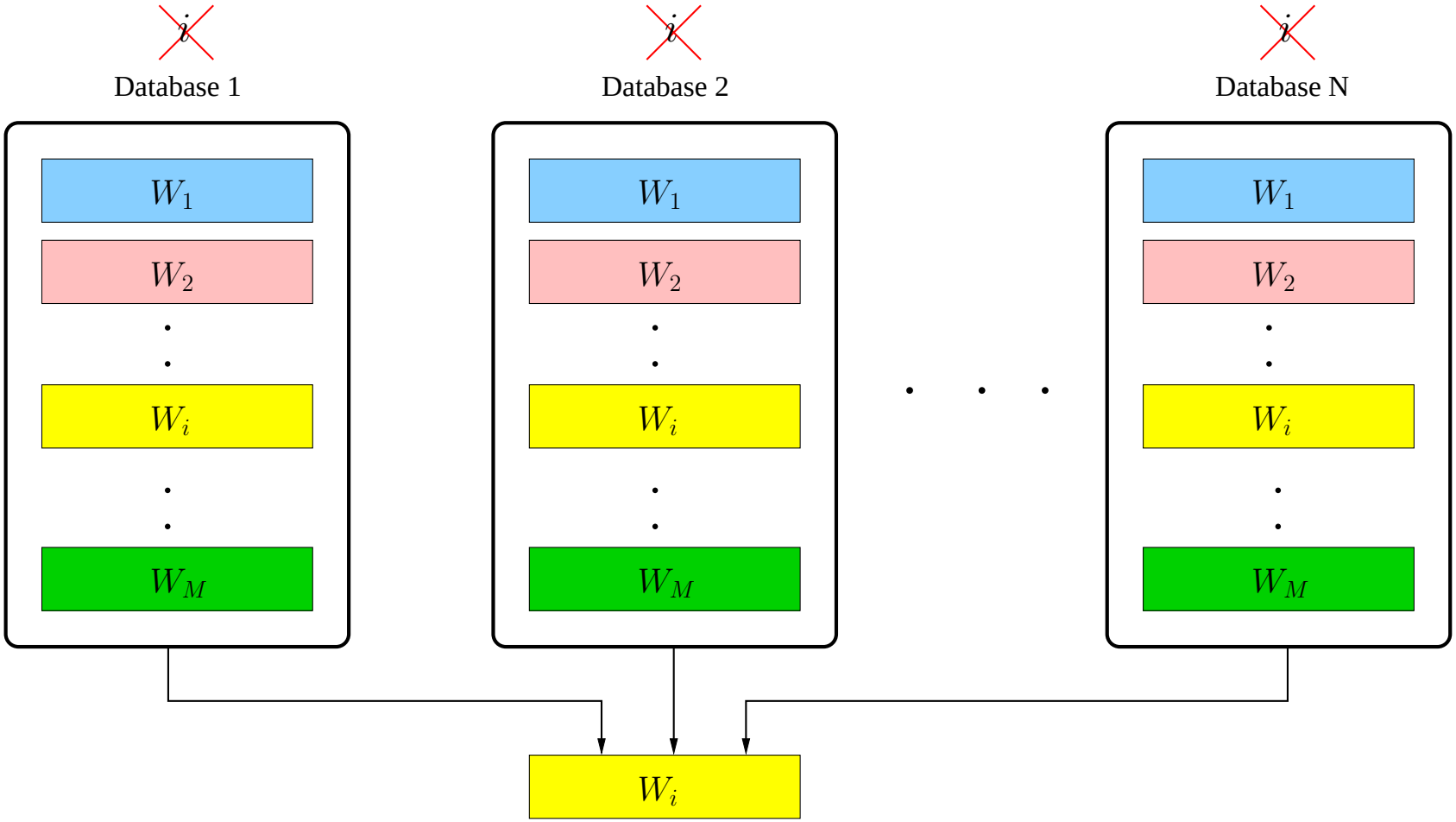
$$H(A_{\mathcal{K}}^{[m]} | Q_{\mathcal{K}}^{[m]}, W_{\mathcal{S}}) = \sum_{n \in \mathcal{K}} H(A_n^{[m]} | Q_n^{[m]}, W_{\mathcal{S}})$$

- **Lemma: Interference lower bound lemma**
  - Lemma does not change due to the distributed storage code.

- **Lemma: Induction lemma**

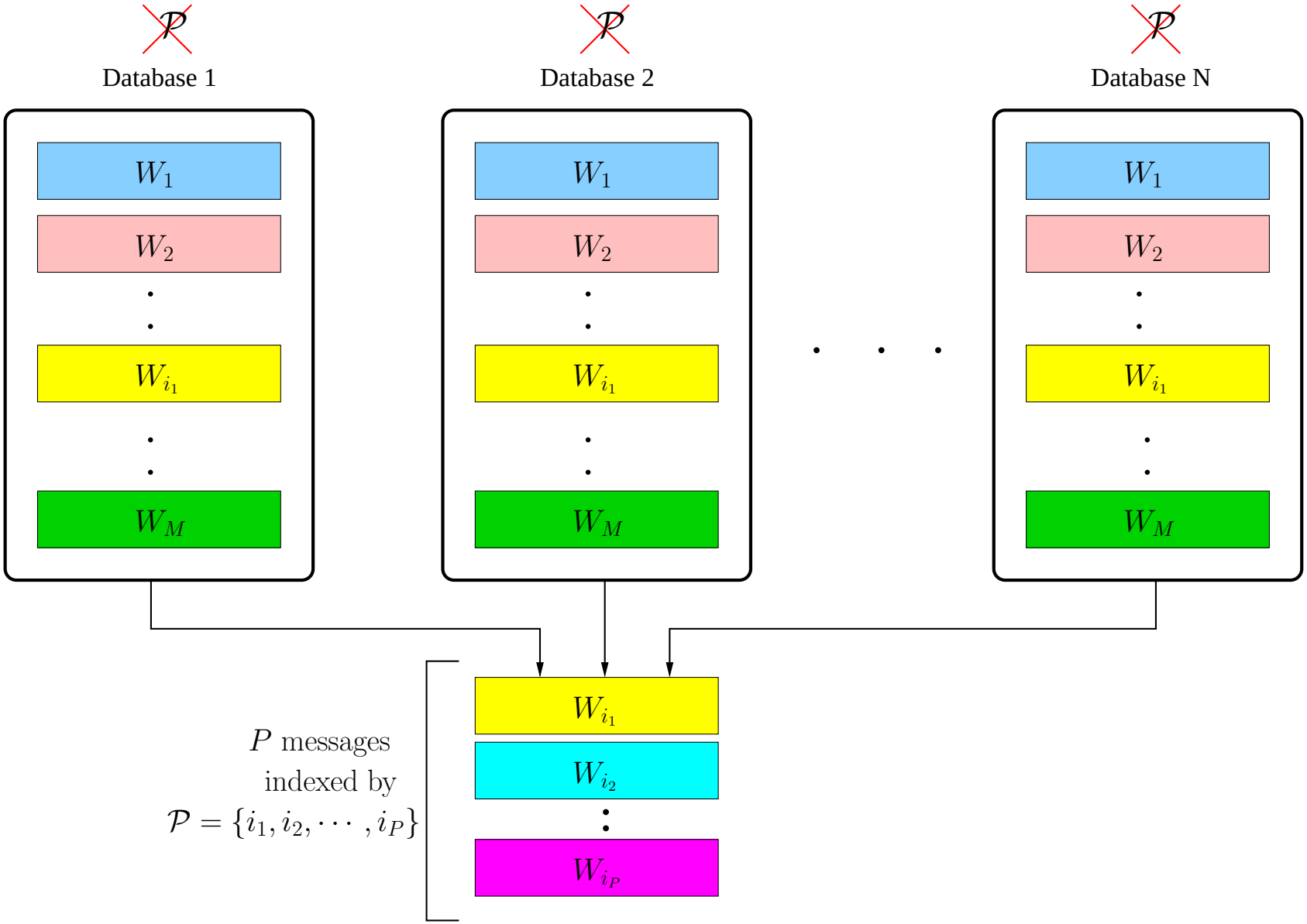
$$I(W_{m:M}; Q_{1:N}^{[m-1]}, A_{1:N}^{[m-1]} | W_{1:m-1})$$
$$\geq \frac{K}{N} I(W_{m+1:M}; Q_{1:N}^{[m]}, A_{1:N}^{[m]} | W_{1:m}) + \frac{KL}{N} - \frac{o(L)}{N}$$

Multi-Message PIR (MPIR) [Banawan-Ulukus][8]



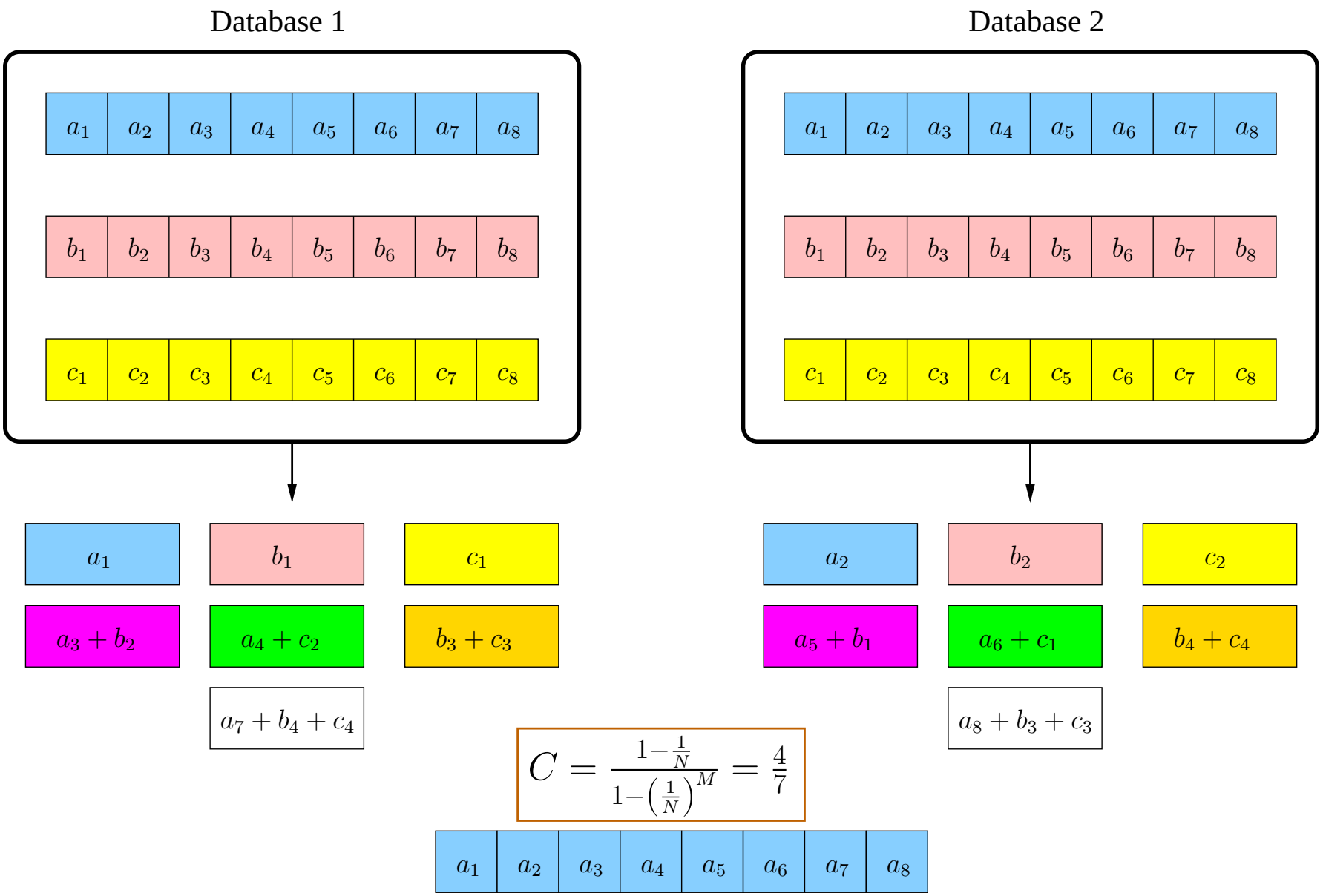
[8] K. Banawan and S. Ulukus, Multi-Message Private Information Retrieval: Capacity Results and Near-Optimal Schemes, IEEE Trans. on Information Theory, to appear. Available at arXiv:1702.01739.

Multi-Message PIR (MPIR) [Banawan-Ulukus][8]

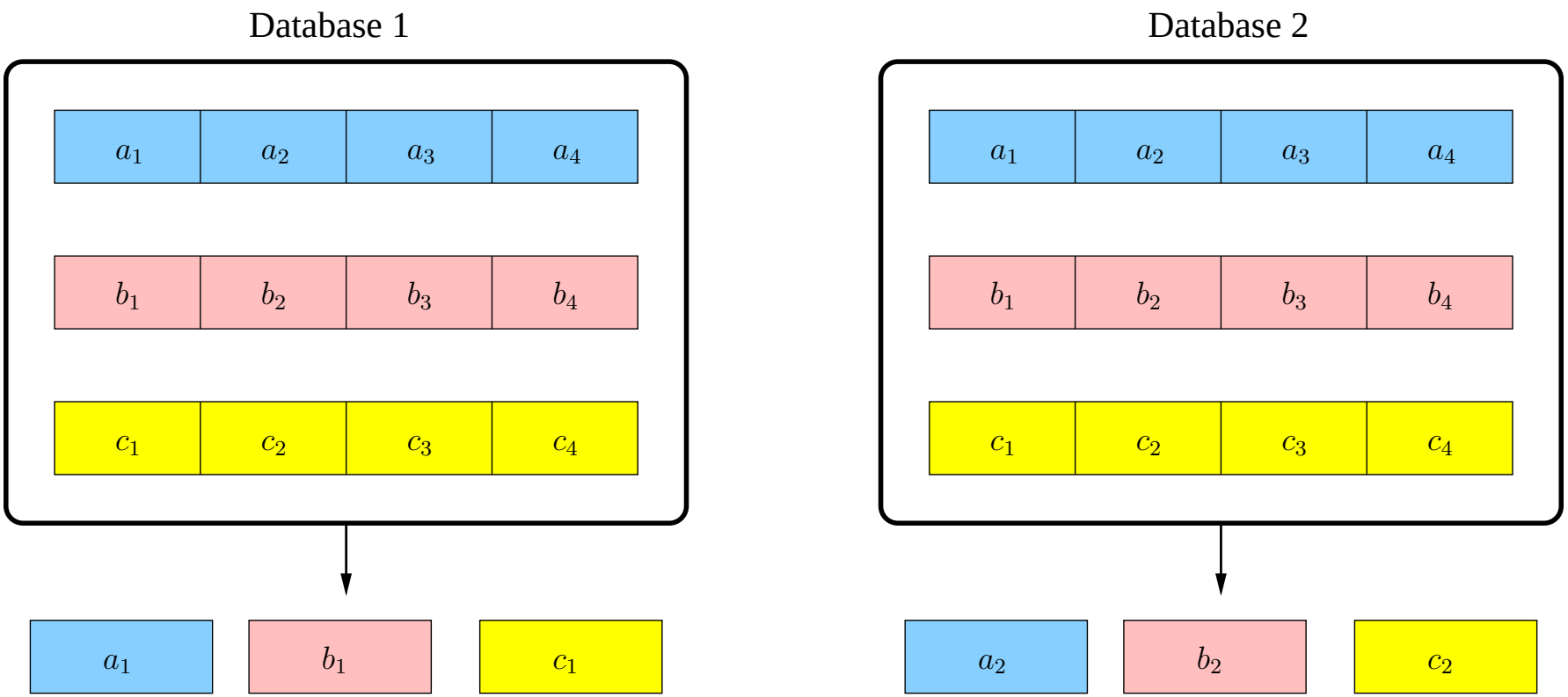


[8] K. Banawan and S. Ulukus, Multi-Message Private Information Retrieval: Capacity Results and Near-Optimal Schemes, IEEE Trans. on Information Theory, to appear. Available at arXiv:1702.01739.

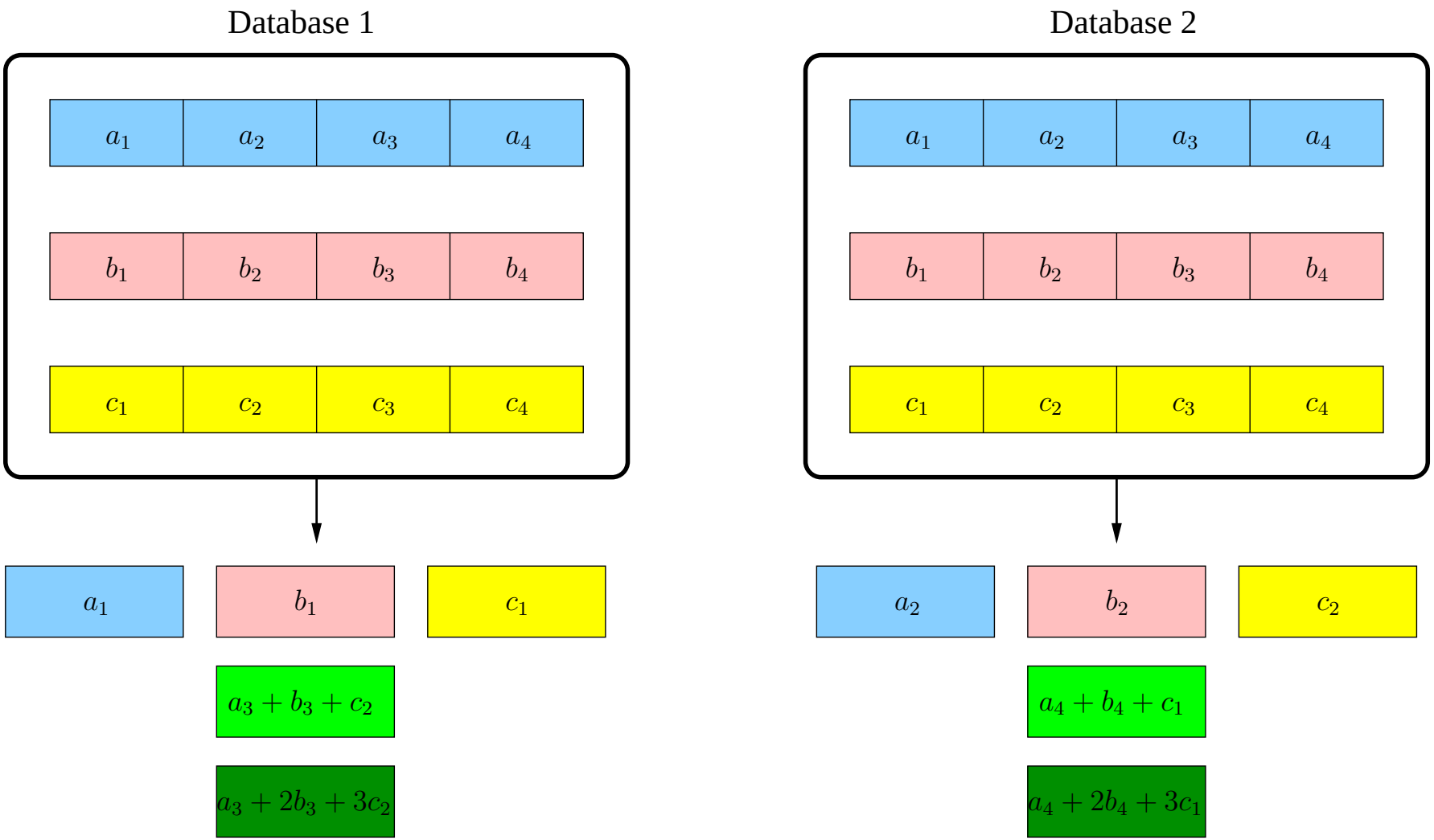
Example $M = 3, N = 2$: Private Retrieval Rate



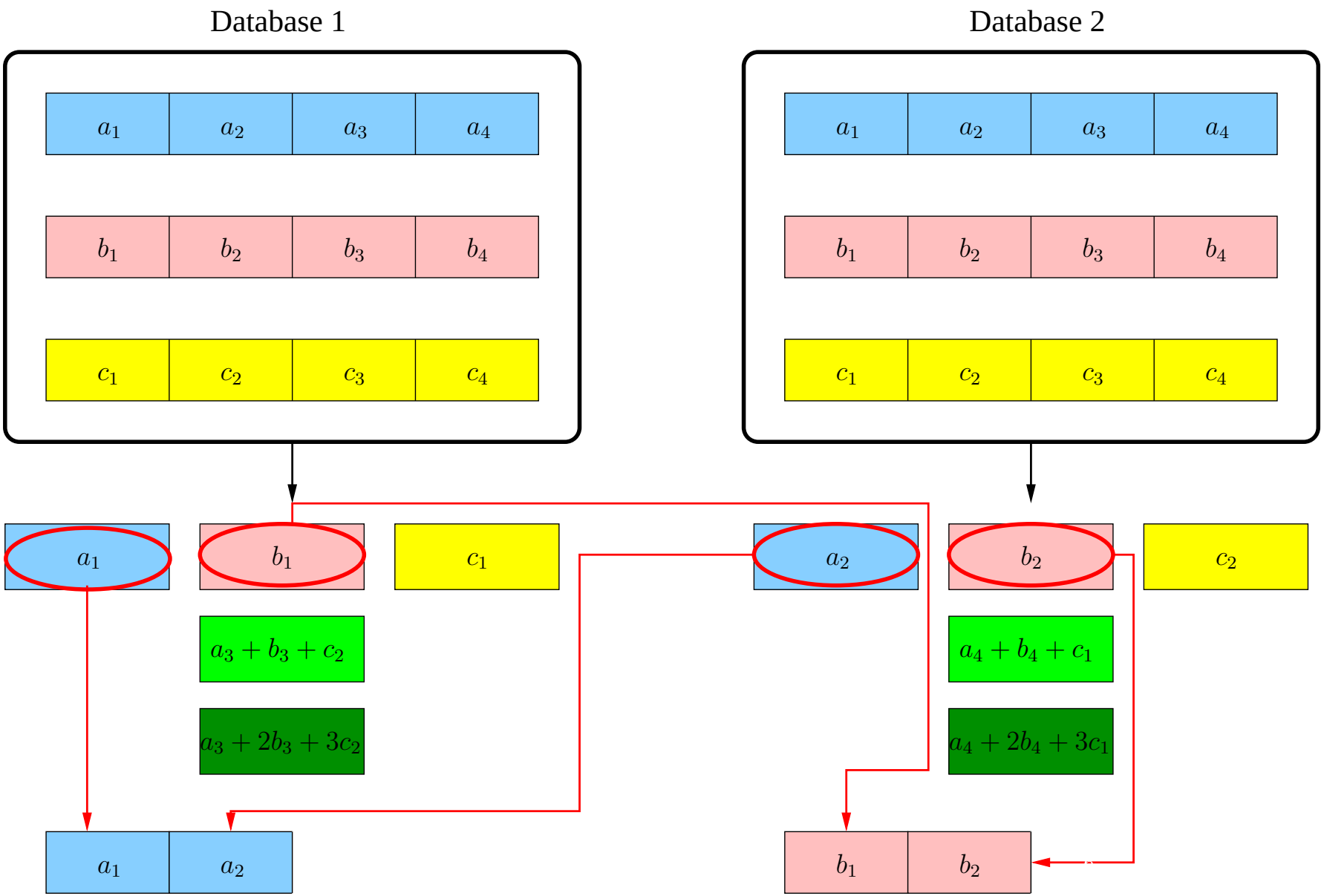
Example $M = 3, N = 2, P = 2$: Joint Retrieval ($P \geq \frac{M}{2}$)



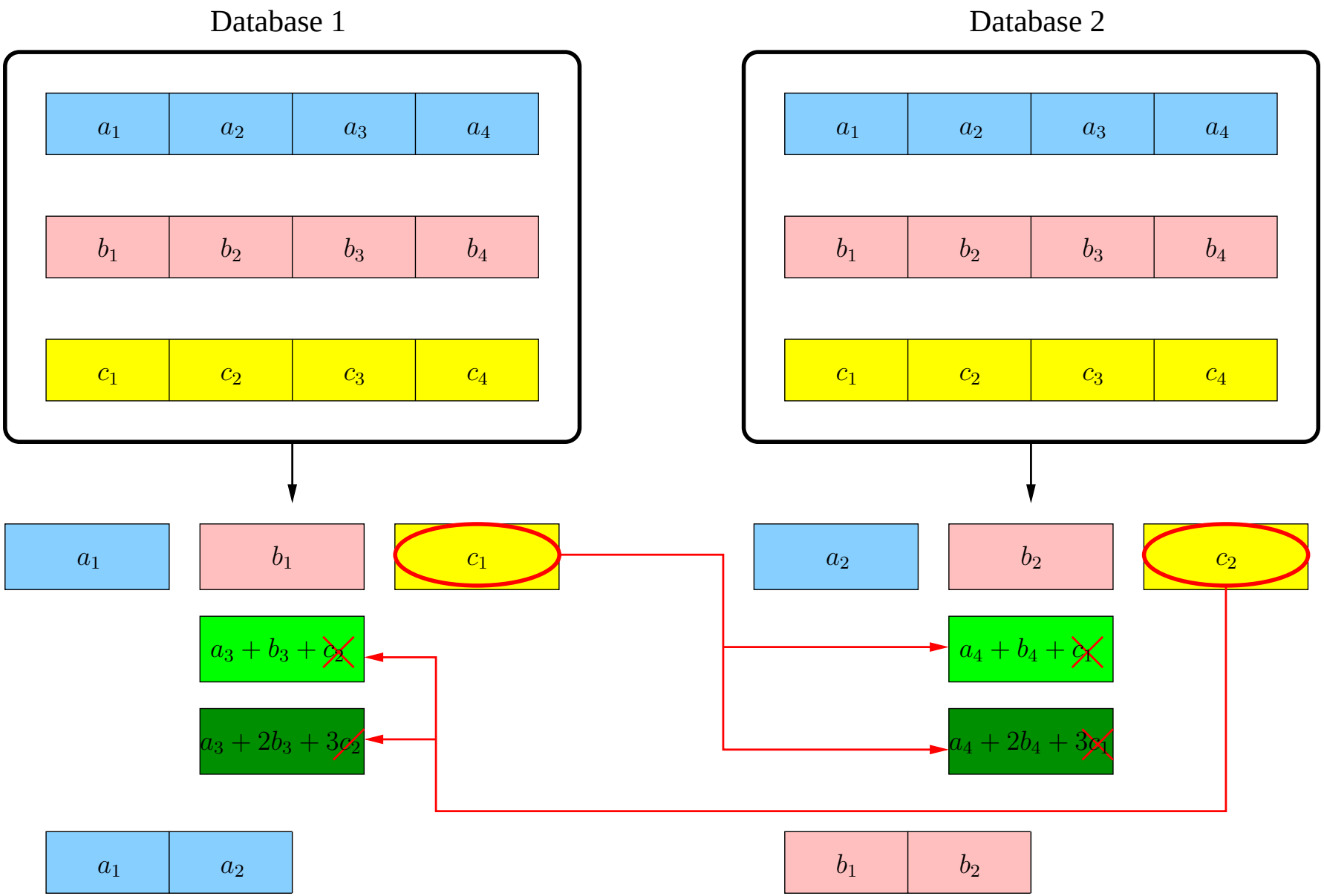
Example $M = 3, N = 2, P = 2$: Joint Retrieval ($P \geq \frac{M}{2}$)



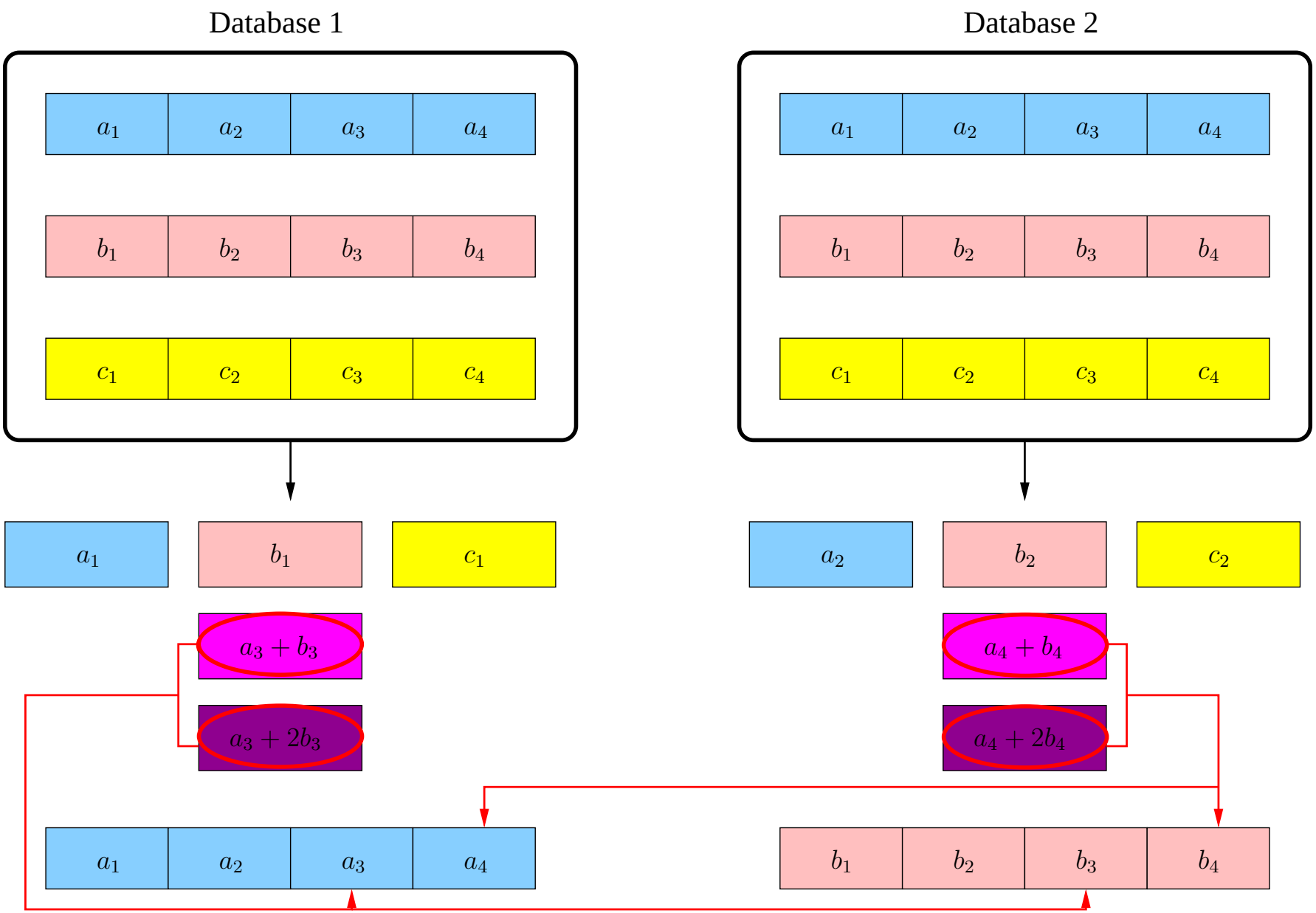
Example $M = 3, N = 2, P = 2$: Joint Decoding ($P \geq \frac{M}{2}$)



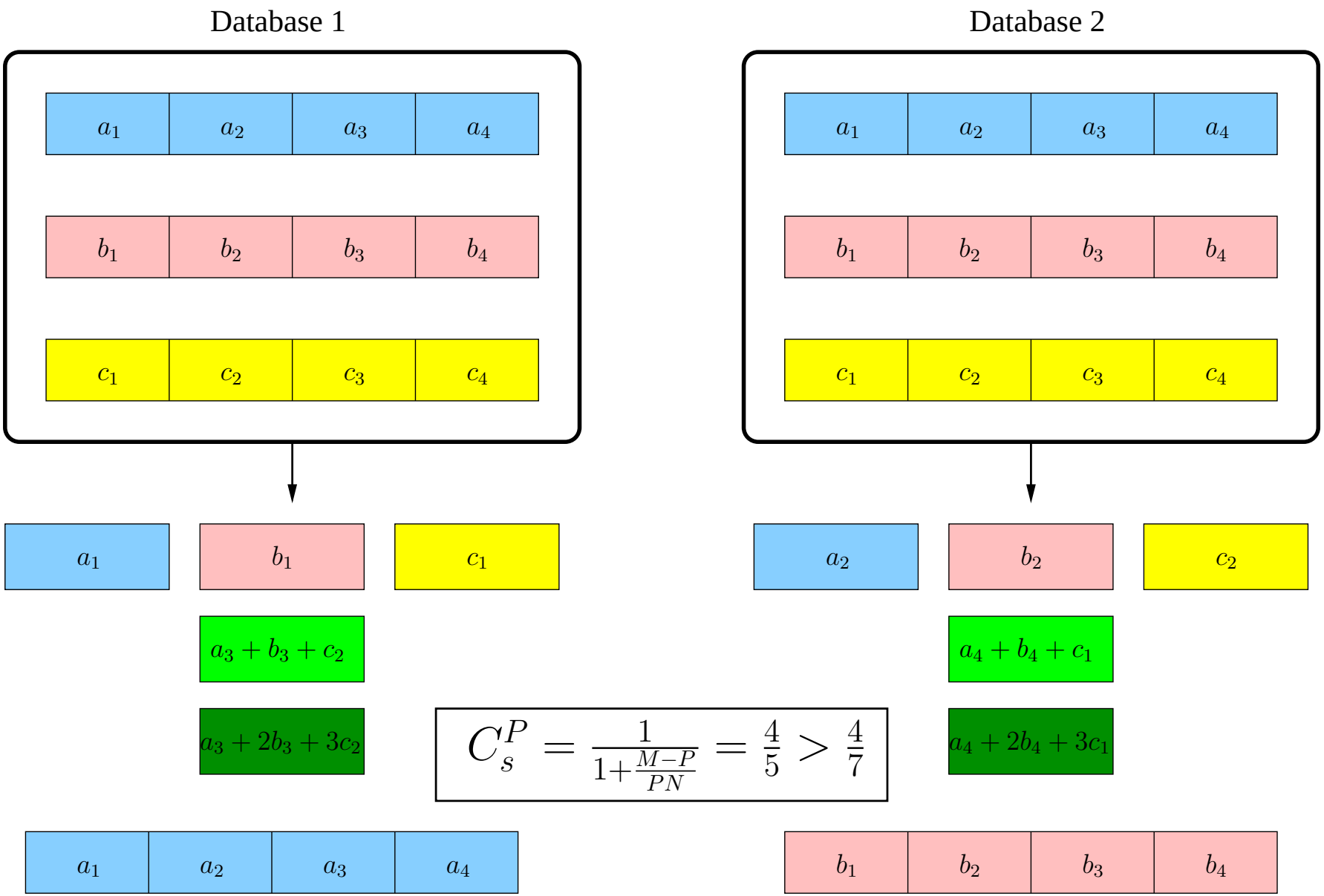
Example $M = 3, N = 2, P = 2$: Joint Decoding ($P \geq \frac{M}{2}$)



Example $M = 3, N = 2, P = 2$: Joint Decoding ($P \geq \frac{M}{2}$)



Example $M = 3, N = 2, P = 2$: Achievable Sum-Rate ($P \geq \frac{M}{2}$)



Example: $M = 5, N = 2, P = 1$ (Classical PIR)

		Database 1	Database 2
rd. 1	stg 1	a_1, b_1, c_1, d_1, e_1	a_2, b_2, c_2, d_2, e_2
round 2	stage 1	$a_3 + b_2$ $a_4 + c_2$ $a_5 + d_2$ $a_6 + e_2$ $b_3 + c_3$ $b_4 + d_3$ $b_5 + e_3$ $c_4 + d_4$ $c_5 + e_4$ $d_5 + e_5$	$a_7 + b_1$ $a_8 + c_1$ $a_9 + d_1$ $a_{10} + e_1$ $b_6 + c_6$ $b_7 + d_6$ $b_8 + e_6$ $c_7 + d_7$ $c_8 + e_7$ $d_8 + e_8$
round 3	stage 1	$a_{11} + b_6 + c_6$ $a_{12} + b_7 + d_6$ $a_{13} + b_8 + e_6$ $a_{14} + c_7 + d_7$ $a_{15} + c_8 + e_7$ $a_{16} + d_8 + e_8$ $b_9 + c_9 + d_9$ $b_{10} + c_{10} + e_9$ $b_{11} + d_{10} + e_{10}$ $c_{11} + d_{11} + e_{11}$	$a_{17} + b_3 + c_3$ $a_{18} + b_4 + d_3$ $a_{19} + b_5 + e_3$ $a_{20} + c_4 + d_4$ $a_{21} + c_5 + e_4$ $a_{22} + d_5 + e_5$ $b_{12} + c_{12} + d_{12}$ $b_{13} + c_{13} + e_{12}$ $b_{14} + d_{13} + e_{13}$ $c_{14} + d_{14} + e_{14}$
round 4	stage 1	$a_{23} + b_{12} + c_{12} + d_{12}$ $a_{24} + b_{13} + c_{13} + e_{12}$ $a_{25} + b_{14} + d_{13} + e_{13}$ $a_{26} + c_{14} + d_{14} + e_{14}$ $b_{15} + c_{15} + d_{15} + e_{15}$	$a_{27} + b_9 + c_9 + d_9$ $a_{28} + b_{10} + c_{10} + e_9$ $a_{29} + b_{11} + d_{10} + e_{10}$ $a_{30} + c_{11} + d_{11} + e_{11}$ $b_{16} + c_{16} + d_{16} + e_{16}$
rd. 5	stg 1	$a_{31} + b_{16} + c_{16} + d_{16} + e_{16}$	$a_{32} + b_{15} + c_{15} + d_{15} + e_{15}$

Example for $P \leq \frac{M}{2}$: $M = 5, N = 2, P = 2$

		Database 1	Database 2
round 1	stg 1	a_1, b_1, c_1, d_1, e_1	a_6, b_6, c_6, d_6, e_6
	stg 2	a_2, b_2, c_2, d_2, e_2	a_7, b_7, c_7, d_7, e_7
	stg 3	a_3, b_3, c_3, d_3, e_3	a_8, b_8, c_8, d_8, e_8
	stg 4	a_4, b_4, c_4, d_4, e_4	a_9, b_9, c_9, d_9, e_9
	stg 5	a_5, b_5, c_5, d_5, e_5	$a_{10}, b_{10}, c_{10}, d_{10}, e_{10}$
round 2	stage 1	$a_{11} + b_6$	$a_{18} + b_1$
		$a_{12} + c_6$	$a_{19} + c_1$
		$a_{13} + d_6$	$a_{20} + d_1$
		$a_{14} + e_6$	$a_{21} + e_1$
		$b_{11} + c_7$	$b_{18} + c_2$
		$b_{12} + d_7$	$b_{19} + d_2$
		$b_{13} + e_7$	$b_{20} + e_2$
		$c_{11} + d_{11}$	$c_{15} + d_{15}$
		$c_{12} + e_{11}$	$c_{16} + e_{15}$
		$d_{12} + e_{12}$	$d_{16} + e_{16}$
	stage 2	$a_6 + b_{14}$	$a_1 + b_{21}$
		$a_{15} + c_8$	$a_{22} + c_3$
		$a_{16} + d_8$	$a_{23} + d_3$
		$a_{17} + e_8$	$a_{24} + e_3$
		$b_{15} + c_9$	$b_{22} + c_4$
		$b_{16} + d_9$	$b_{23} + d_4$
		$b_{17} + e_9$	$b_{24} + e_4$
		$c_{13} + d_{13}$	$c_{17} + d_{17}$
		$c_{14} + e_{13}$	$c_{18} + e_{17}$
		$d_{14} + e_{14}$	$d_{18} + e_{18}$
round 3	stage 1	$a_{25} + b_7 + c_{10}$	$a_2 + b_{29} + c_5$
		$a_7 + b_{25} + d_{10}$	$a_{30} + b_2 + d_5$
		$a_{26} + b_8 + e_{10}$	$a_3 + b_{30} + e_5$
		$a_{27} + c_{15} + d_{15}$	$a_{31} + c_{11} + d_{11}$
		$a_{28} + c_{16} + e_{15}$	$a_{32} + c_{12} + e_{11}$
		$a_{29} + d_{16} + e_{16}$	$a_{33} + d_{12} + e_{12}$
		$b_{26} + c_{17} + d_{17}$	$b_{31} + c_{13} + d_{13}$
		$b_{27} + c_{18} + e_{17}$	$b_{32} + c_{14} + e_{13}$
		$b_{28} + d_{18} + e_{18}$	$b_{33} + d_{14} + e_{14}$
		$c_{19} + d_{19} + e_{19}$	$c_{20} + d_{20} + e_{20}$
rd. 5	stg 1	$a_8 + b_{34} + c_{20} + d_{20} + e_{20}$	$a_{34} + b_3 + c_{19} + d_{19} + e_{19}$

Example for $P \leq \frac{M}{2}$: $M = 5, N = 2, P = 2$

		Database 1	Database 2
round 1	stg 1	a_1, b_1, c_1, d_1, e_1	a_6, b_6, c_6, d_6, e_6
	stg 2	a_2, b_2, c_2, d_2, e_2	a_7, b_7, c_7, d_7, e_7
	stg 3	a_3, b_3, c_3, d_3, e_3	a_8, b_8, c_8, d_8, e_8
	stg 4	a_4, b_4, c_4, d_4, e_4	a_9, b_9, c_9, d_9, e_9
	stg 5	a_5, b_5, c_5, d_5, e_5	$a_{10}, b_{10}, c_{10}, d_{10}, e_{10}$
round 2	stage 1	$a_{11} + b_6$	$a_{18} + b_1$
		$a_{12} + c_6$	$a_{19} + c_1$
		$a_{13} + d_6$	$a_{20} + d_1$
		$a_{14} + e_6$	$a_{21} + e_1$
		$b_{11} + c_7$	$b_{18} + c_2$
		$b_{12} + d_7$	$b_{19} + d_2$
		$b_{13} + e_7$	$b_{20} + e_2$
		$c_{11} + d_{11}$	$c_{15} + d_{15}$
		$c_{12} + e_{11}$	$c_{16} + e_{15}$
		$d_{12} + e_{12}$	$d_{16} + e_{16}$
	stage 2	$a_6 + b_{14}$	$a_1 + b_{21}$
		$a_{15} + c_8$	$a_{22} + c_3$
		$a_{16} + d_8$	$a_{23} + d_3$
		$a_{17} + e_8$	$a_{24} + e_3$
		$b_{15} + c_9$	$b_{22} + c_4$
		$b_{16} + d_9$	$b_{23} + d_4$
		$b_{17} + e_9$	$b_{24} + e_4$
		$c_{13} + d_{13}$	$c_{17} + d_{17}$
		$c_{14} + e_{13}$	$c_{18} + e_{17}$
		$d_{14} + e_{14}$	$d_{18} + e_{18}$
round 3	stage 1	$a_{25} + b_7 + c_{10}$	$a_2 + b_{29} + c_5$
		$a_7 + b_{25} + d_{10}$	$a_{30} + b_2 + d_5$
		$a_{26} + b_8 + e_{10}$	$a_3 + b_{30} + e_5$
		$a_{27} + c_{15} + d_{15}$	$a_{31} + c_{11} + d_{11}$
		$a_{28} + c_{16} + e_{15}$	$a_{32} + c_{12} + e_{11}$
		$a_{29} + d_{16} + e_{16}$	$a_{33} + d_{12} + e_{12}$
		$b_{26} + c_{17} + d_{17}$	$b_{31} + c_{13} + d_{13}$
		$b_{27} + c_{18} + e_{17}$	$b_{32} + c_{14} + e_{13}$
		$b_{28} + d_{18} + e_{18}$	$b_{33} + d_{14} + e_{14}$
		$\boxed{\mathbf{c_{19} + d_{19} + e_{19}}}$	$\boxed{\mathbf{c_{20} + d_{20} + e_{20}}}$
rd. 5	stg 1	$a_8 + b_{34} + \boxed{\mathbf{c_{20} + d_{20} + e_{20}}}$	$a_{34} + b_3 + \boxed{\mathbf{c_{19} + d_{19} + e_{19}}}$

Example for $P \leq \frac{M}{2}$: $M = 5, N = 2, P = 2$

		Database 1	Database 2
round 1	stg 1	a_1, b_1, c_1, d_1, e_1	a_6, b_6, c_6, d_6, e_6
	stg 2	a_2, b_2, c_2, d_2, e_2	a_7, b_7, c_7, d_7, e_7
	stg 3	a_3, b_3, c_3, d_3, e_3	a_8, b_8, c_8, d_8, e_8
	stg 4	a_4, b_4, c_4, d_4, e_4	a_9, b_9, c_9, d_9, e_9
	stg 5	$a_5, b_5, \mathbf{c_5}, \mathbf{d_5}, \mathbf{e_5}$	$a_{10}, b_{10}, \mathbf{c_{10}}, \mathbf{d_{10}}, \mathbf{e_{10}}$
round 2	stage 1	$a_{11} + b_6$	$a_{18} + b_1$
		$a_{12} + c_6$	$a_{19} + c_1$
		$a_{13} + d_6$	$a_{20} + d_1$
		$a_{14} + e_6$	$a_{21} + e_1$
		$b_{11} + c_7$	$b_{18} + c_2$
		$b_{12} + d_7$	$b_{19} + d_2$
		$b_{13} + e_7$	$b_{20} + e_2$
		$c_{11} + d_{11}$	$c_{15} + d_{15}$
		$c_{12} + e_{11}$	$c_{16} + e_{15}$
		$d_{12} + e_{12}$	$d_{16} + e_{16}$
	stage 2	$a_6 + b_{14}$	$a_1 + b_{21}$
		$a_{15} + c_8$	$a_{22} + c_3$
		$a_{16} + d_8$	$a_{23} + d_3$
		$a_{17} + e_8$	$a_{24} + e_3$
		$b_{15} + c_9$	$b_{22} + c_4$
		$b_{16} + d_9$	$b_{23} + d_4$
		$b_{17} + e_9$	$b_{24} + e_4$
		$c_{13} + d_{13}$	$c_{17} + d_{17}$
		$c_{14} + e_{13}$	$c_{18} + e_{17}$
		$d_{14} + e_{14}$	$d_{18} + e_{18}$
round 3	stage 1	$a_{25} + b_7 + \mathbf{c_{10}}$	$a_2 + b_{29} + \mathbf{c_5}$
		$a_7 + b_{25} + \mathbf{d_{10}}$	$a_{30} + b_2 + \mathbf{d_5}$
		$a_{26} + b_8 + \mathbf{e_{10}}$	$a_3 + b_{30} + \mathbf{e_5}$
		$a_{27} + c_{15} + d_{15}$	$a_{31} + c_{11} + d_{11}$
		$a_{28} + c_{16} + e_{15}$	$a_{32} + c_{12} + e_{11}$
		$a_{29} + d_{16} + e_{16}$	$a_{33} + d_{12} + e_{12}$
		$b_{26} + c_{17} + d_{17}$	$b_{31} + c_{13} + d_{13}$
		$b_{27} + c_{18} + e_{17}$	$b_{32} + c_{14} + e_{13}$
		$b_{28} + d_{18} + e_{18}$	$b_{33} + d_{14} + e_{14}$
		$c_{19} + d_{19} + e_{19}$	$c_{20} + d_{20} + e_{20}$
rd. 5	stg 1	$a_8 + b_{34} + c_{20} + d_{20} + e_{20}$	$a_{34} + b_3 + c_{19} + d_{19} + e_{19}$

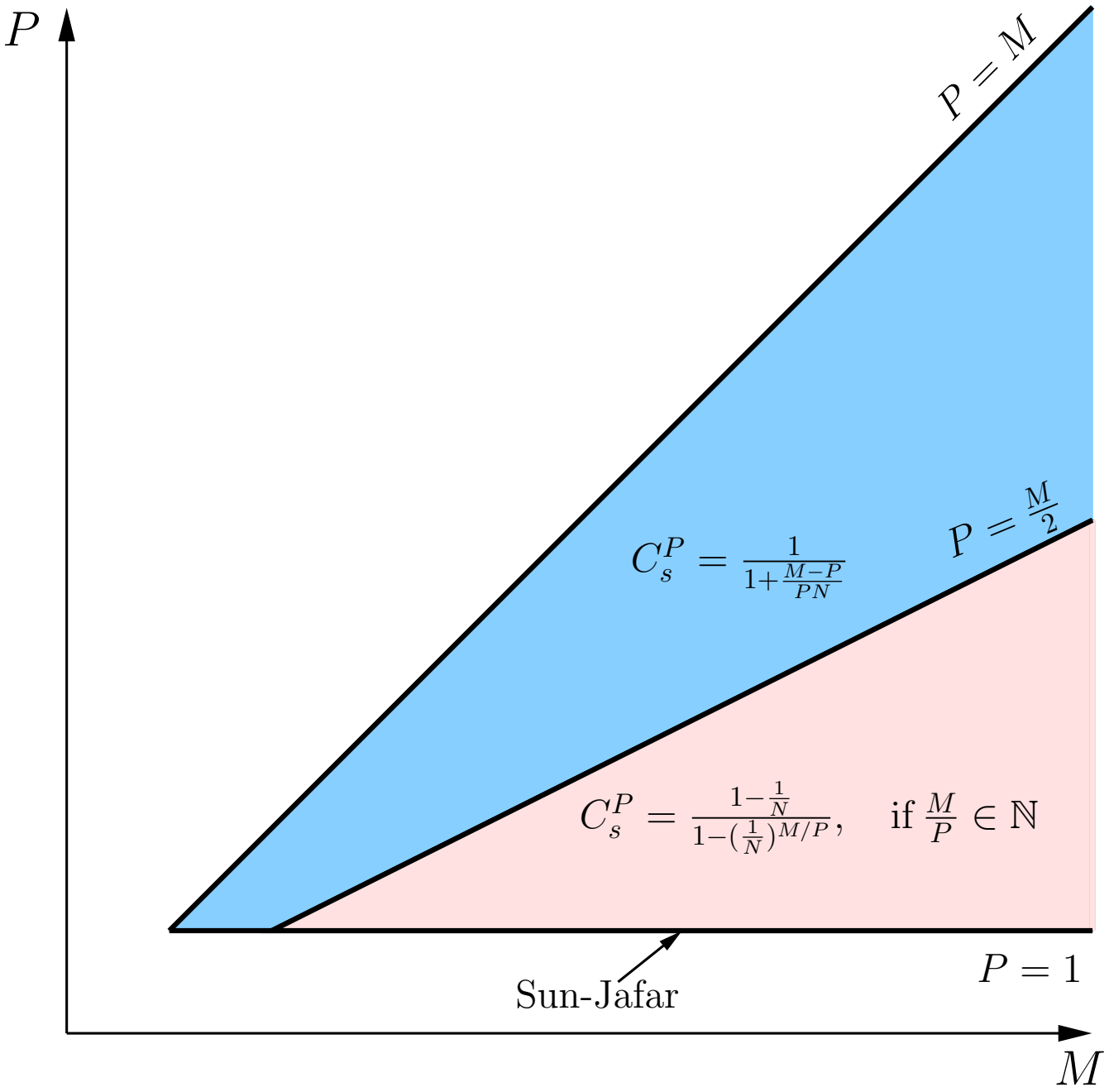
Example for $P \leq \frac{M}{2}$: $M = 5, N = 2, P = 2$

		Database 1	Database 2
round 1	stg 1	a_1, b_1, c_1, d_1, e_1	a_6, b_6, c_6, d_6, e_6
	stg 2	a_2, b_2, c_2, d_2, e_2	a_7, b_7, c_7, d_7, e_7
	stg 3	a_3, b_3, c_3, d_3, e_3	a_8, b_8, c_8, d_8, e_8
	stg 4	a_4, b_4, c_4, d_4, e_4	a_9, b_9, c_9, d_9, e_9
	stg 5	a_5, b_5, c_5, d_5, e_5	$a_{10}, b_{10}, c_{10}, d_{10}, e_{10}$
round 2	stage 1	$a_{11} + b_6$ $a_{12} + c_6$ $a_{13} + d_6$ $a_{14} + e_6$ $b_{11} + c_7$ $b_{12} + d_7$ $b_{13} + e_7$ $\mathbf{c_{11} + d_{11}}$ $\mathbf{c_{12} + e_{11}}$ $\mathbf{d_{12} + e_{12}}$	$a_{18} + b_1$ $a_{19} + c_1$ $a_{20} + d_1$ $a_{21} + e_1$ $b_{18} + c_2$ $b_{19} + d_2$ $b_{20} + e_2$ $\mathbf{c_{15} + d_{15}}$ $\mathbf{c_{16} + e_{15}}$ $\mathbf{d_{16} + e_{16}}$
	stage 2	$a_6 + b_{14}$ $a_{15} + c_8$ $a_{16} + d_8$ $a_{17} + e_8$ $b_{15} + c_9$ $b_{16} + d_9$ $b_{17} + e_9$ $\mathbf{c_{13} + d_{13}}$ $\mathbf{c_{14} + e_{13}}$ $\mathbf{d_{14} + e_{14}}$	$a_1 + b_{21}$ $a_{22} + c_3$ $a_{23} + d_3$ $a_{24} + e_3$ $b_{22} + c_4$ $b_{23} + d_4$ $b_{24} + e_4$ $\mathbf{c_{17} + d_{17}}$ $\mathbf{c_{18} + e_{17}}$ $\mathbf{d_{18} + e_{18}}$
round 3	stage 1	$a_{25} + b_7 + c_{10}$ $a_7 + b_{25} + d_{10}$ $a_{26} + b_8 + e_{10}$ $a_{27} + \mathbf{c_{15} + d_{15}}$ $a_{28} + \mathbf{c_{16} + e_{15}}$ $a_{29} + \mathbf{d_{16} + e_{16}}$ $b_{26} + \mathbf{c_{17} + d_{17}}$ $b_{27} + \mathbf{c_{18} + e_{17}}$ $b_{28} + \mathbf{d_{18} + e_{18}}$ $c_{19} + d_{19} + e_{19}$	$a_2 + b_{29} + c_5$ $a_{30} + b_2 + d_5$ $a_3 + b_{30} + e_5$ $a_{31} + \mathbf{c_{11} + d_{11}}$ $a_{32} + \mathbf{c_{12} + e_{11}}$ $a_{33} + \mathbf{d_{12} + e_{12}}$ $b_{31} + \mathbf{c_{13} + d_{13}}$ $b_{32} + \mathbf{c_{14} + e_{13}}$ $b_{33} + \mathbf{d_{14} + e_{14}}$ $c_{20} + d_{20} + e_{20}$
rd. 5	stg 1	$a_8 + b_{34} + c_{20} + d_{20} + e_{20}$	$a_{34} + b_3 + c_{19} + d_{19} + e_{19}$

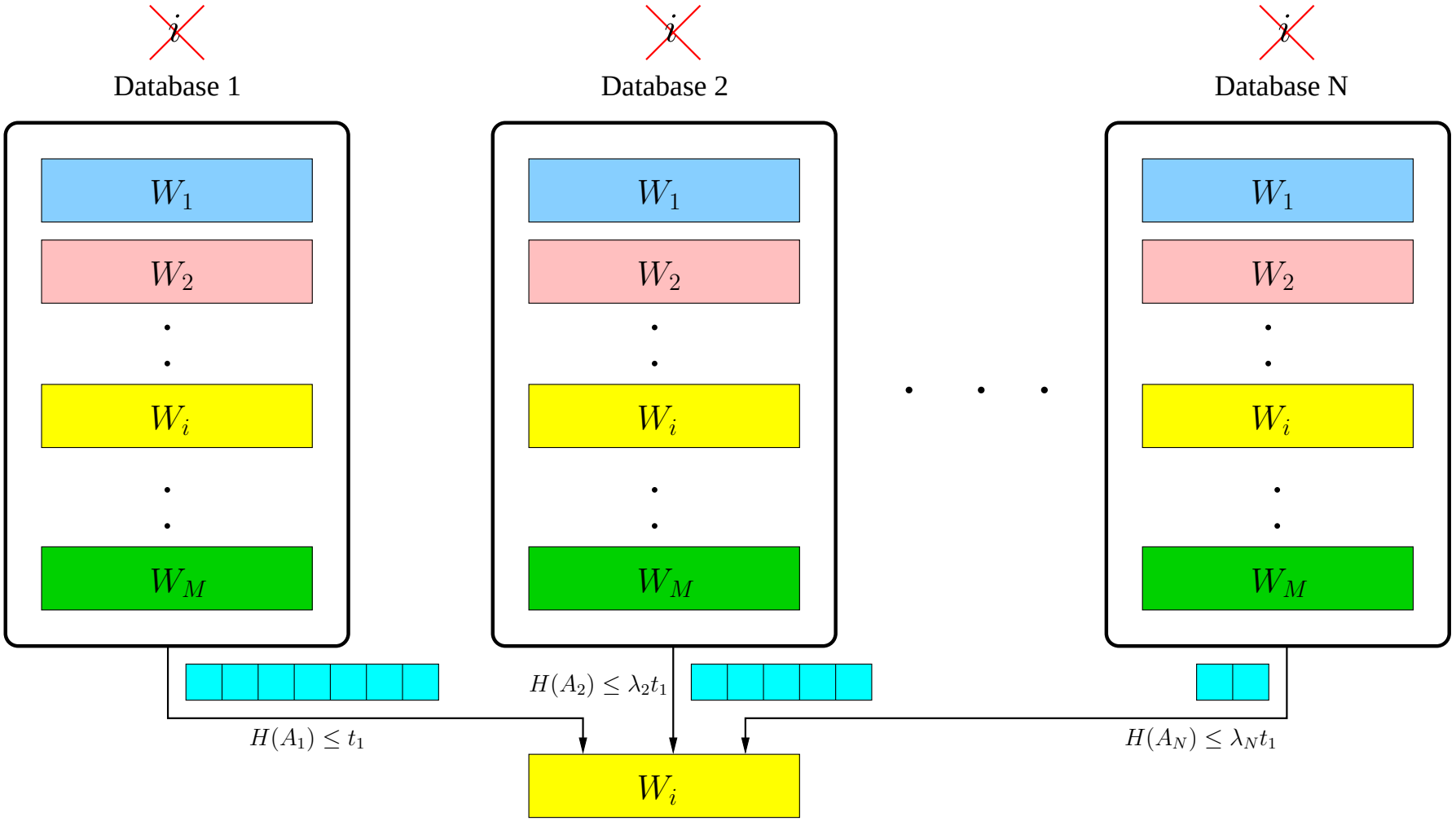
Example for $P \leq \frac{M}{2}$: $M = 5, N = 2, P = 2$

		Database 1	Database 2
round 1	stg 1	$a_1, b_1, \mathbf{c_1}, \mathbf{d_1}, \mathbf{e_1}$	$a_6, b_6, \mathbf{c_6}, \mathbf{d_6}, \mathbf{e_6}$
	stg 2	$a_2, b_2, \mathbf{c_2}, \mathbf{d_2}, \mathbf{e_2}$	$a_7, b_7, \mathbf{c_7}, \mathbf{d_7}, \mathbf{e_7}$
	stg 3	$a_3, b_3, \mathbf{c_3}, \mathbf{d_3}, \mathbf{e_3}$	$a_8, b_8, \mathbf{c_8}, \mathbf{d_8}, \mathbf{e_8}$
	stg 4	$a_4, b_4, \mathbf{c_4}, \mathbf{d_4}, \mathbf{e_4}$	$a_9, b_9, \mathbf{c_9}, \mathbf{d_9}, \mathbf{e_9}$
	stg 5	a_5, b_5, c_5, d_5, e_5	$a_{10}, b_{10}, c_{10}, d_{10}, e_{10}$
round 2	stage 1	$a_{11} + b_6$	$a_{18} + b_1$
		$a_{12} + \mathbf{c_6}$	$a_{19} + \mathbf{c_1}$
		$a_{13} + \mathbf{d_6}$	$a_{20} + \mathbf{d_1}$
		$a_{14} + \mathbf{e_6}$	$a_{21} + \mathbf{e_1}$
		$b_{11} + \mathbf{c_7}$	$b_{18} + \mathbf{c_2}$
		$b_{12} + \mathbf{d_7}$	$b_{19} + \mathbf{d_2}$
		$b_{13} + \mathbf{e_7}$	$b_{20} + \mathbf{e_2}$
		$c_{11} + d_{11}$	$c_{15} + d_{15}$
		$c_{12} + e_{11}$	$c_{16} + e_{15}$
		$d_{12} + e_{12}$	$d_{16} + e_{16}$
	stage 2	$a_6 + b_{14}$	$a_1 + b_{21}$
		$a_{15} + \mathbf{c_8}$	$a_{22} + \mathbf{c_3}$
		$a_{16} + \mathbf{d_8}$	$a_{23} + \mathbf{d_3}$
		$a_{17} + \mathbf{e_8}$	$a_{24} + \mathbf{e_3}$
		$b_{15} + \mathbf{c_9}$	$b_{22} + \mathbf{c_4}$
		$b_{16} + \mathbf{d_9}$	$b_{23} + \mathbf{d_4}$
		$b_{17} + \mathbf{e_9}$	$b_{24} + \mathbf{e_4}$
		$c_{13} + d_{13}$	$c_{17} + d_{17}$
		$c_{14} + e_{13}$	$c_{18} + e_{17}$
		$d_{14} + e_{14}$	$d_{18} + e_{18}$
round 3	stage 1	$a_{25} + b_7 + c_{10}$	$a_2 + b_{29} + c_5$
		$a_7 + b_{25} + d_{10}$	$a_{30} + b_2 + d_5$
		$a_{26} + b_8 + e_{10}$	$a_3 + b_{30} + e_5$
		$a_{27} + c_{15} + d_{15}$	$a_{31} + c_{11} + d_{11}$
		$a_{28} + c_{16} + e_{15}$	$a_{32} + c_{12} + e_{11}$
		$a_{29} + d_{16} + e_{16}$	$a_{33} + d_{12} + e_{12}$
		$b_{26} + c_{17} + d_{17}$	$b_{31} + c_{13} + d_{13}$
		$b_{27} + c_{18} + e_{17}$	$b_{32} + c_{14} + e_{13}$
		$b_{28} + d_{18} + e_{18}$	$b_{33} + d_{14} + e_{14}$
		$c_{19} + d_{19} + e_{19}$	$c_{20} + d_{20} + e_{20}$
rd. 5	stg 1	$a_8 + b_{34} + c_{20} + d_{20} + e_{20}$	$a_{34} + b_3 + c_{19} + d_{19} + e_{19}$

Multi-Message PIR Capacity



PIR Under Asymmetric Traffic Constraints [Banawan-Ulukus][9]



[9] K. Banawan and S. Ulukus. Asymmetry hurts: Private information retrieval under asymmetric traffic constraints. IEEE Trans. on Info. Theory, 2018. Available at arXiv:1801.03079.

Introducing Asymmtery: PIR with Arbitrary Message Length [Sun-Jafar][10]

Database 1	Database 2
a_1 b_1 c_1	a_2 b_2 c_2
$a_3 + b_2$ $a_4 + c_2$ $b_3 + c_3$	$a_5 + b_1$ $a_6 + c_1$ $b_4 + c_4$
$a_7 + b_4 + c_4$	$a_8 + b_3 + c_3$

$$\boxed{R = \frac{1 - \frac{1}{N}}{1 - (\frac{1}{N})^M} = \frac{1 - \frac{1}{2}}{1 - (\frac{1}{2})^3} = \frac{8}{14} = \frac{4}{7}}$$

[10] H. Sun and S. Jafar. Optimal download cost of private information retrieval for arbitrary message length. 2016. Available at arXiv:1610.03048.

Introducing Asymmtery: PIR with Arbitrary Message Length [Sun-Jafar][10]

Database 1	Database 2
a_1 b_1 c_1	
	$a_2 + b_1$ $a_3 + c_1$ $b_2 + c_2$
$a_4 + b_2 + c_2$	

$$R = \frac{1 - \frac{1}{N}}{1 - (\frac{1}{N})^M} = \frac{1 - \frac{1}{2}}{1 - (\frac{1}{2})^3} = \frac{4}{7}$$

[10] H. Sun and S. Jafar. Optimal download cost of private information retrieval for arbitrary message length. 2016. Available at arXiv:1610.03048.

Introducing Asymmtery: PIR with Arbitrary Message Length [Sun-Jafar][10]

Database 1	Database 2
a_1 b_1 c_1	
	$a_2 + b_1$ $a_3 + c_1$ $b_2 + c_2$
$a_4 + b_2 + c_2$	

$$R = \frac{1 - \frac{1}{N}}{1 - (\frac{1}{N})^M} = \frac{1 - \frac{1}{2}}{1 - (\frac{1}{2})^3} = \frac{4}{7}$$

[10] H. Sun and S. A. Jafar. Optimal download cost of private information retrieval for arbitrary message length. IEEE Trans. on Info. Forensics and Security, 12(12):2920–2932, Dec 2017.

Introducing Asymmtery: PIR with Arbitrary Message Length [Sun-Jafar][10]

Database 1	Database 2
a_1 b_1 c_1	
	$a_2 + b_1$ $a_3 + c_1$ $b_2 + c_2$
$a_4 + b_2 + c_2$	

$$R = \frac{1 - \frac{1}{N}}{1 - (\frac{1}{N})^M} = \frac{1 - \frac{1}{2}}{1 - (\frac{1}{2})^3} = \frac{4}{7}$$

- Ratio between traffic is 4 : 3.

[10] H. Sun and S. A. Jafar. Optimal download cost of private information retrieval for arbitrary message length. IEEE Trans. on Info. Forensics and Security, 12(12):2920–2932, Dec 2017.

Introducing Asymmtery: PIR with Arbitrary Message Length [Sun-Jafar][10]

Database 1	Database 2
a_1 b_1 c_1	
	$a_2 + b_1$ $a_3 + c_1$ $b_2 + c_2$
$a_4 + b_2 + c_2$	

$$R = \frac{1 - \frac{1}{N}}{1 - (\frac{1}{N})^M} = \frac{1 - \frac{1}{2}}{1 - (\frac{1}{2})^3} = \frac{4}{7}$$

- Ratio between traffic is $4 : 3$.
- **How to achieve a general traffic ratio $\lambda_1 : \lambda_2$?**

[10] H. Sun and S. A. Jafar. Optimal download cost of private information retrieval for arbitrary message length. IEEE Trans. on Info. Forensics and Security, 12(12):2920–2932, Dec 2017.

Introducing Asymmtery: PIR with Arbitrary Message Length [Sun-Jafar][10]

Database 1	Database 2
a_1 b_1 c_1	
	$a_2 + b_1$ $a_3 + c_1$ $b_2 + c_2$
$a_4 + b_2 + c_2$	

$$R = \frac{1 - \frac{1}{N}}{1 - (\frac{1}{N})^M} = \frac{1 - \frac{1}{2}}{1 - (\frac{1}{2})^3} = \frac{4}{7}$$

- Ratio between traffic is $4 : 3$.
- **How to achieve a general traffic ratio $\lambda_1 : \lambda_2$?**
- A reason for different traffic ratios: different link capacities.

[10] H. Sun and S. A. Jafar. Optimal download cost of private information retrieval for arbitrary message length. IEEE Trans. on Info. Forensics and Security, 12(12):2920–2932, Dec 2017.

Asymmetric Traffic Constraints

- The nth database responds with a t_n-length answer string.
- The lengths of the answer strings are different.

$$t_n = \lambda_n t_1, \quad 1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_N \text{ (wlog)}$$

- The ratios between the traffic are

$$1 : \lambda_2 : \lambda_3 : \cdots : \lambda_N$$

- Traffic ratio of the nth database

$$\tau_n = \frac{\lambda_n}{\sum_{j=1}^{N} \lambda_j}$$

Asymmetric Traffic Constraints

- The nth database responds with a t_n-length answer string.
- The lengths of the answer strings are different.

$$t_n = \lambda_n t_1, \quad 1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_N \text{ (wlog)}$$

- The ratios between the traffic are

$$1 : \lambda_2 : \lambda_3 : \cdots : \lambda_N$$

- Traffic ratio of the nth database

$$\tau_n = \frac{\lambda_n}{\sum_{j=1}^{N} \lambda_j}$$

- $(\lambda_1, \cdots, \lambda_N)$ have a one-to-one relationship with $(\tau_1, \cdots, \tau_N)$.

Asymmetric Traffic Constraints

- The nth database responds with a t_n-length answer string.
- The lengths of the answer strings are different.

$$t_n = \lambda_n t_1, \quad 1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_N \text{ (wlog)}$$

- The ratios between the traffic are

$$1 : \lambda_2 : \lambda_3 : \cdots : \lambda_N$$

- Traffic ratio of the nth database

$$\tau_n = \frac{\lambda_n}{\sum_{j=1}^{N} \lambda_j}$$

- $(\lambda_1, \cdots, \lambda_N)$ have a one-to-one relationship with $(\tau_1, \cdots, \tau_N)$.
- **Does asymmetry hurt the retrieval rate?**

Converse Proof: New Tools and Lemmas

- **Lemma: Interference lower bound lemma**
  - No change as it deals with the length of the entire downloaded answers.

- **Lemma: Induction lemma**
  For all $m \in \{2, \ldots, M\}$ and for an arbitrary $n_{m-1} \in \{1, \cdots, N\}$,

$$I\left(W_{m:M}; Q_{1:N}^{[m-1]}, A_{1:N}^{[m-1]} | W_{1:m-1}\right)$$

$$\geq \underbrace{\frac{1}{n_{m-1}} I\left(W_{m+1:M}; Q_{1:N}^{[m]}, A_{1:N}^{[m]} | W_{1:m}\right)}_{\substack{n_{m-1} \text{ is the number of databases applying} \\ \text{symmetric schemes for messages } W_{m-1:M}}} + \underbrace{\frac{1}{n_{m-1}}\left(L - t_1 \sum_{n=n_{m-1}+1}^{N} \lambda_n\right)}_{\substack{\text{remaining answers are bounded trivially} \\ \text{by the length of the answer string } t_n}} - \frac{o(L)}{n_{m-1}}$$

Converse Proof: Tightest Upper Bound

- **Ordering terms and minimizing over $n_i \in \{1, \cdots, N\}$ leads to**

$$\boxed{C(\boldsymbol{\tau}) \leq \bar{C}(\boldsymbol{\tau}) = \min_{n_i \in \{1,\cdots,N\}} \frac{1 + \frac{\sum_{n=n_1+1}^{N} \tau_n}{n_1} + \frac{\sum_{n=n_2+1}^{N} \tau_n}{n_1 n_2} + \cdots + \frac{\sum_{n=n_{M-1}+1}^{N} \tau_n}{n_1 \cdots n_{M-1}}}{1 + \frac{1}{n_1} + \frac{1}{n_1 n_2} + \cdots + \frac{1}{n_1 \cdots n_{M-1}}}}$$

- If $n_1 = n_2 = \cdots = n_{M-1} = N$ (Sun-Jafar bound)

$$\boxed{C(\boldsymbol{\tau}) \leq \frac{1}{1 + \frac{1}{N} + \cdots + \frac{1}{N^{M-1}}}}$$

- If $\boldsymbol{\tau} = (1, 0, \cdots, 0)$, and we pick $n_1 = n_2 = \cdots = n_{M-1} = 1$ (trivial bound)

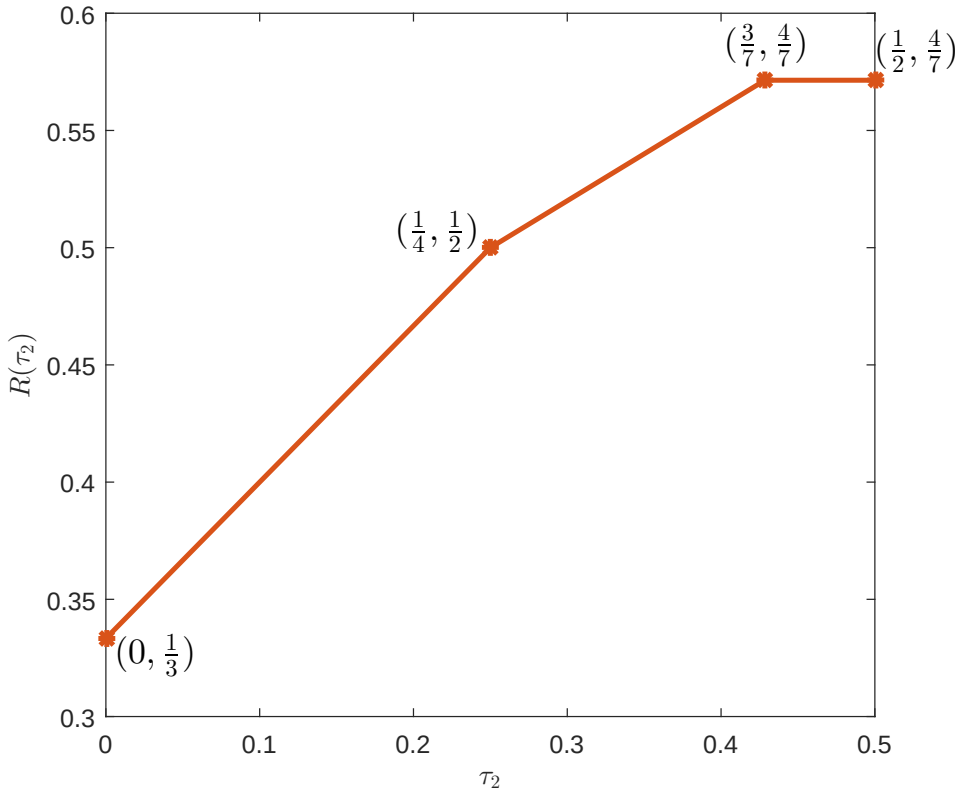
$$\boxed{C(\boldsymbol{\tau}) \leq \frac{1}{M}}$$

Example: $M = 3,\ N = 2$

- Traffic ratio $\tau_2 = \frac{\lambda_2}{\lambda_1+\lambda_2}$.
- **Explicit upper bound:** by minimizing over $n_1,\ n_2$

$$C(\tau_2) \leq \begin{cases} \frac{1}{3} + \frac{2\tau_2}{3}, & 0 \leq \tau_2 \leq \frac{1}{4} \\ \frac{2}{5} + \frac{2\tau_2}{5}, & \frac{1}{4} \leq \tau_2 \leq \frac{3}{7} \\ \frac{4}{7}, & \frac{3}{7} \leq \tau_2 \leq \frac{1}{2} \end{cases}$$

- **Asymmetry hurts:** $\lambda_2 < \frac{3}{4}$ $(\tau_2 < \frac{3}{7})$ incurs capacity loss.



Achievability of Corner Points: $M = 3$, $N = 2$

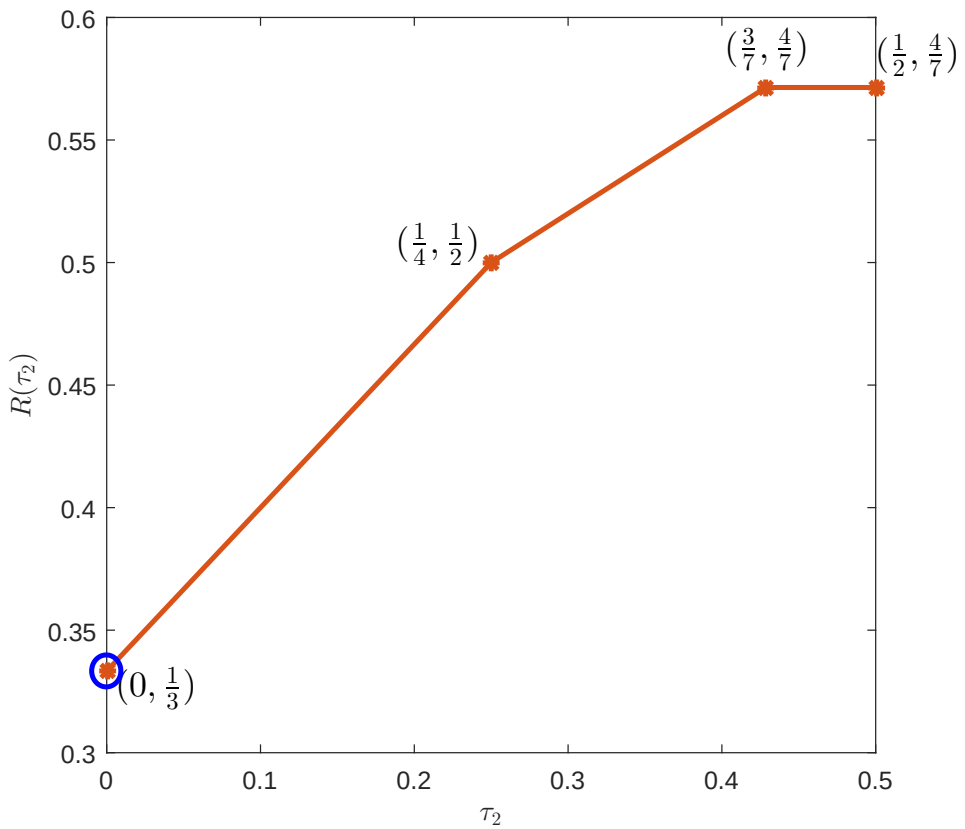
- **The $\tau_2 = 0$ Corner Point:** This achieves $R = \frac{1}{3} = C(0)$.

Database 1	Database 2
a_1, b_1, c_1	

Achievability of Corner Points: $M = 3$, $N = 2$

- **The $\tau_2 = 0$ Corner Point:** This achieves $R = \frac{1}{3} = C(0)$.

Database 1	Database 2
a_1, b_1, c_1	



Achievability of Corner Points: $M = 3$, $N = 2$

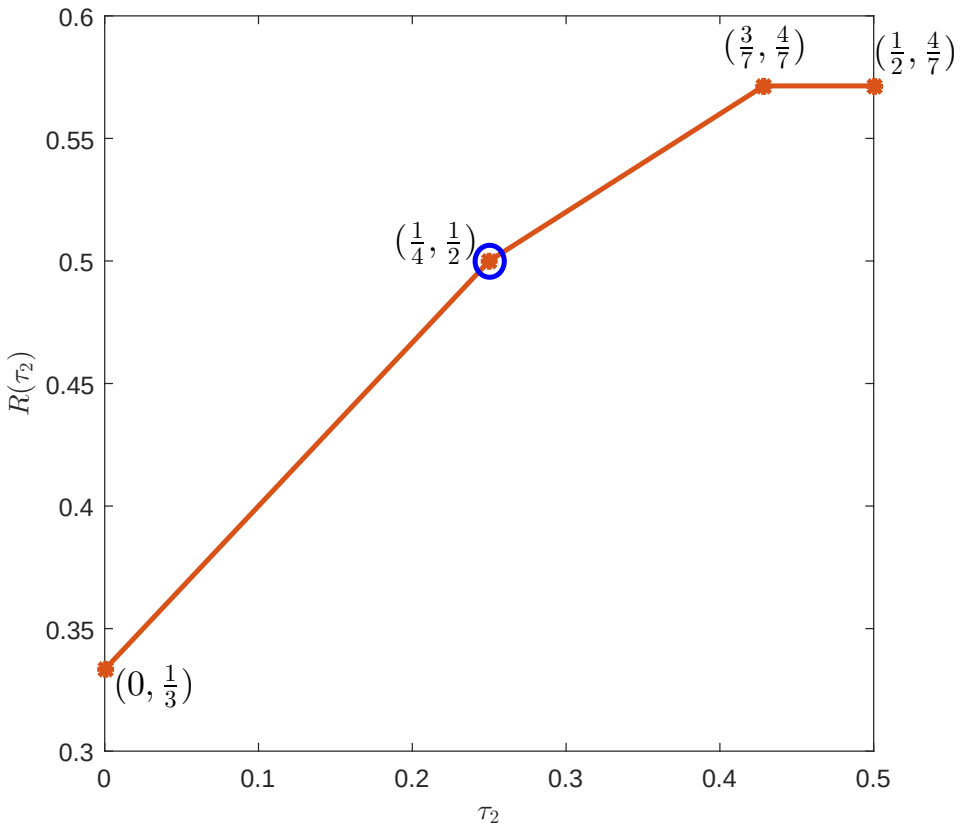
- **The $\tau_2 = \frac{1}{4}$ Corner Point:** This achieves $R = \frac{1}{2} = C(\frac{1}{4})$.

Database 1	Database 2
a_1, b_1, c_1	
	$a_2 + b_1 + c_1$

Achievability of Corner Points: $M = 3$, $N = 2$

- **The $\tau_2 = \frac{1}{4}$ Corner Point:** This achieves $R = \frac{1}{2} = C(\frac{1}{4})$.

Database 1	Database 2
a_1, b_1, c_1	
	$a_2 + b_1 + c_1$



Achievability of Corner Points: $M = 3$, $N = 2$

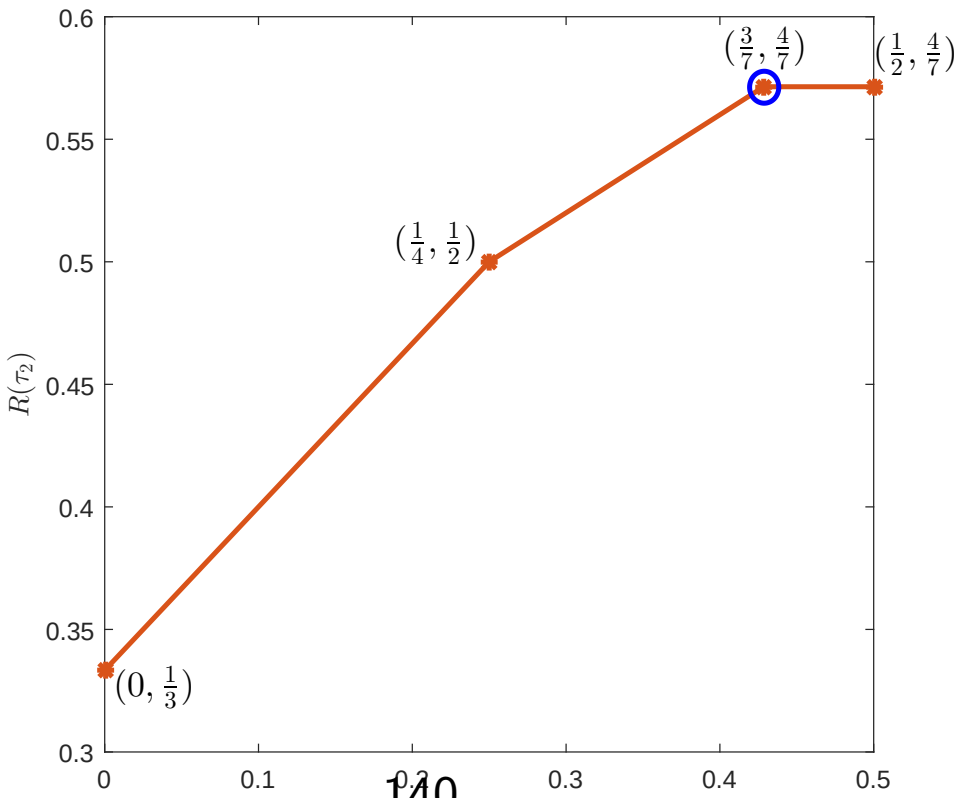
- **The $\tau_2 = \frac{3}{7}$ Corner Point:** This achieves $R = \frac{4}{7} = C(\frac{3}{7})$.

Database 1	Database 2
a_1, b_1, c_1	
	$a_2 + b_1$ $a_3 + c_1$ $b_2 + c_2$
$a_4 + b_2 + c_2$	

Achievability of Corner Points: $M = 3$, $N = 2$

- **The $\tau_2 = \frac{3}{7}$ Corner Point:** This achieves $R = \frac{4}{7} = C(\frac{3}{7})$.

Database 1	Database 2
a_1, b_1, c_1	
	$a_2 + b_1$ $a_3 + c_1$ $b_2 + c_2$
$a_4 + b_2 + c_2$	



Achievability of Corner Points: $M = 3$, $N = 2$

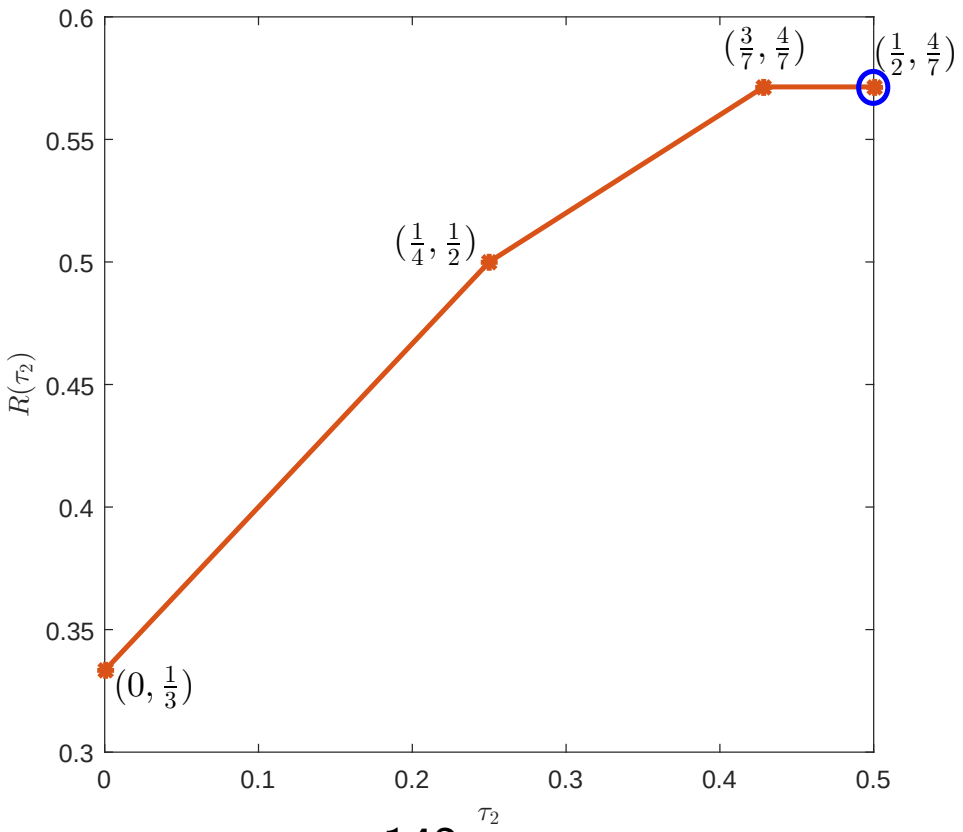
- **The $\tau_2 = \frac{1}{2}$ Corner Point:** Sun-Jafar symmetric scheme.

Database 1	Database 2
a_1, b_1, c_1	a_2, b_2, c_2
$a_3 + b_2$ $a_4 + c_2$ $b_3 + c_3$	$a_5 + b_1$ $a_6 + c_1$ $b_4 + c_4$
$a_7 + b_4 + c_4$	$a_8 + b_3 + c_3$

Achievability of Corner Points: $M = 3$, $N = 2$

- **The $\tau_2 = \frac{1}{2}$ Corner Point:** Sun-Jafar symmetric scheme.

Database 1	Database 2
a_1, b_1, c_1	a_2, b_2, c_2
$a_3 + b_2$ $a_4 + c_2$ $b_3 + c_3$	$a_5 + b_1$ $a_6 + c_1$ $b_4 + c_4$
$a_7 + b_4 + c_4$	$a_8 + b_3 + c_3$



Achievability of Non-Corner Points: $M = 3$, $N = 2$

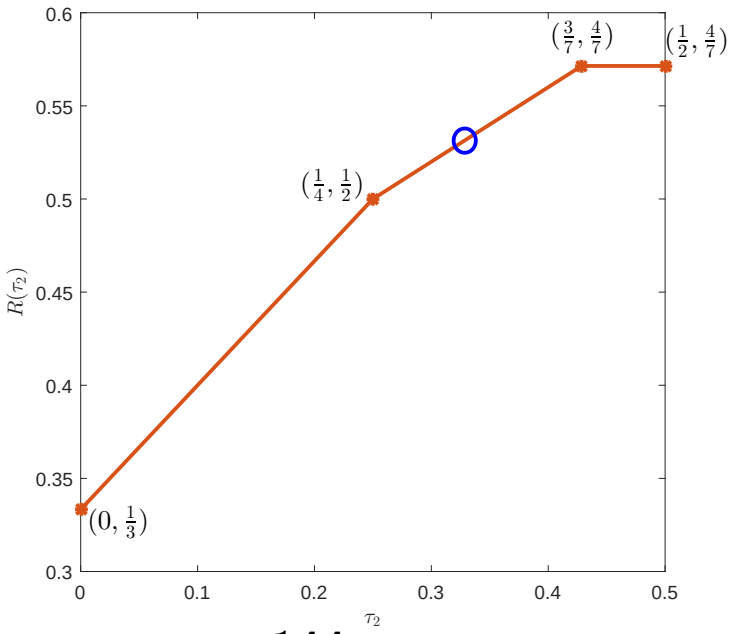
- Upper bound is affine in $\boldsymbol{\tau}$.
- Achievability of any non-corner points is done by time-sharing.
- **Example:** $\tau_2 = \frac{1}{3}$**:** $R(\frac{1}{3}) = \frac{8}{15} = \frac{2}{5} + \frac{2\tau_2}{5} = C(\frac{1}{3})$.

Database 1	Database 2
a_1, b_1, c_1	$a_2 + b_1$ $a_3 + c_1$ $b_2 + c_2$
$a_4 + b_2 + c_2$	
a_5, b_3, c_3	$a_6 + b_3 + c_3$
a_7, b_4, c_4	$a_8 + b_4 + c_4$

Achievability of Non-Corner Points: $M = 3$, $N = 2$

- Upper bound is affine in $\boldsymbol{\tau}$.
- Achievability of any non-corner points is done by time-sharing.
- **Example:** $\tau_2 = \frac{1}{3}$**:** $R(\frac{1}{3}) = \frac{8}{15} = \frac{2}{5} + \frac{2\tau_2}{5} = C(\frac{1}{3})$.

Database 1	Database 2
a_1, b_1, c_1	$a_2 + b_1$ $a_3 + c_1$ $b_2 + c_2$
$a_4 + b_2 + c_2$	
a_5, b_3, c_3	$a_6 + b_3 + c_3$
a_7, b_4, c_4	$a_8 + b_4 + c_4$



Achievability of Non-Corner Points: $M = 3$, $N = 2$

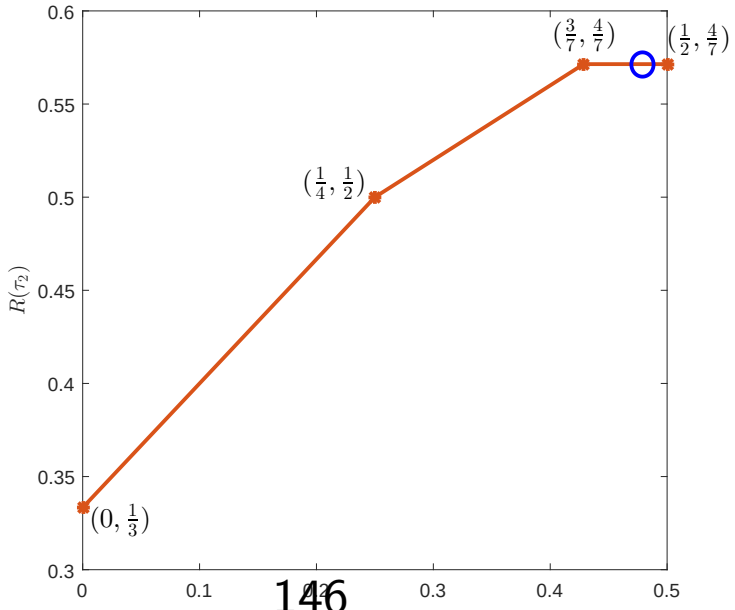
- **Example:** $\tau_2 = \frac{10}{21}$**:** $R(\frac{10}{21}) = \frac{4}{7} = C(\frac{10}{21})$.

Database 1	Database 2
a_1, b_1, c_1	a_2, b_2, c_2
$a_3 + b_2$ $a_4 + c_2$ $b_3 + c_3$	$a_5 + b_1$ $a_6 + c_1$ $b_4 + c_4$
$a_7 + b_4 + c_4$	$a_8 + b_3 + c_3$
a_9, b_5, c_5	$a_{10} + b_5$ $a_{11} + c_5$ $b_6 + c_6$
$a_{12} + b_6 + c_6$	

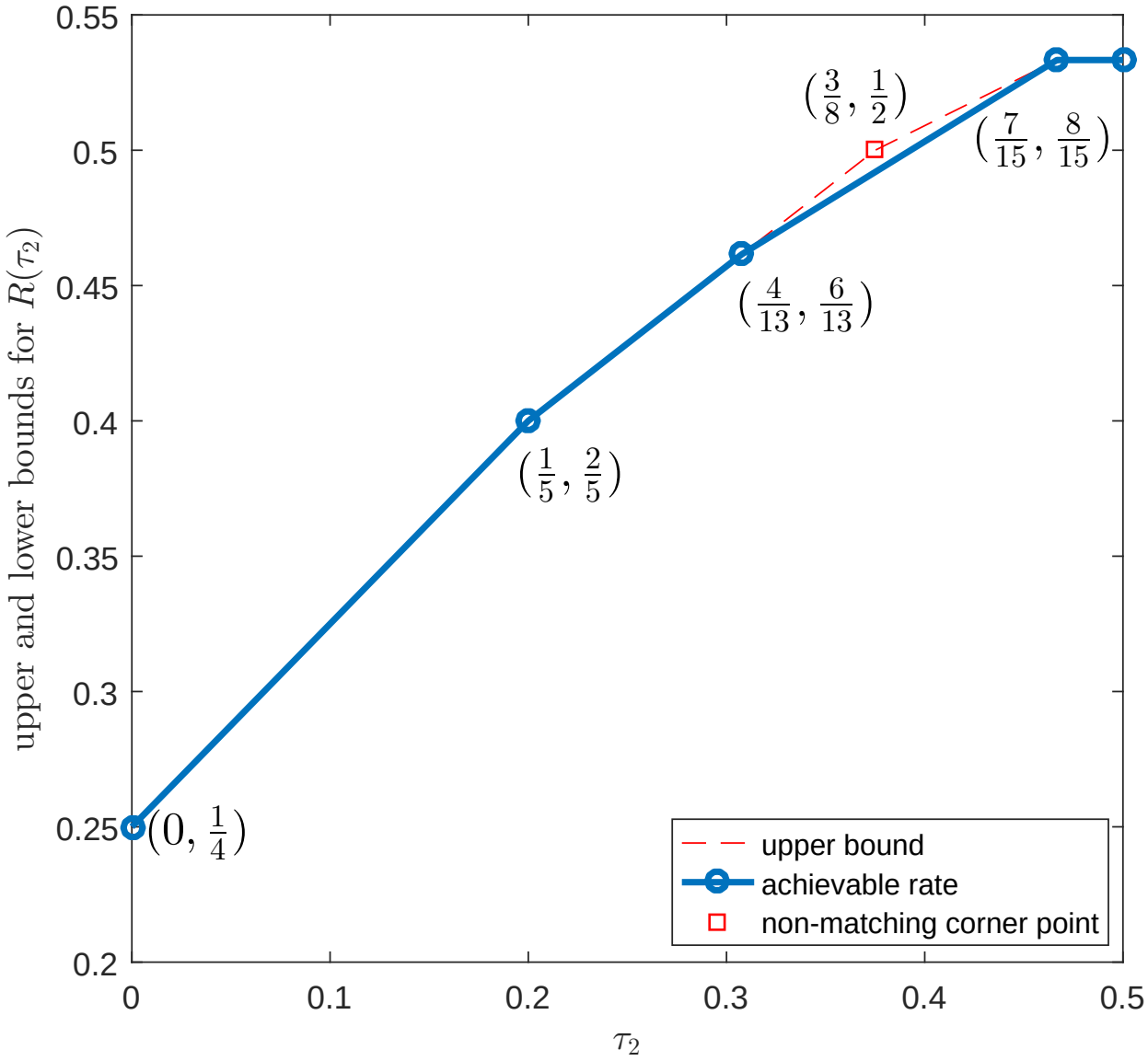
Achievability of Non-Corner Points: $M = 3$, $N = 2$

- **Example:** $\tau_2 = \frac{10}{21}$**:** $R(\frac{10}{21}) = \frac{4}{7} = C(\frac{10}{21})$.

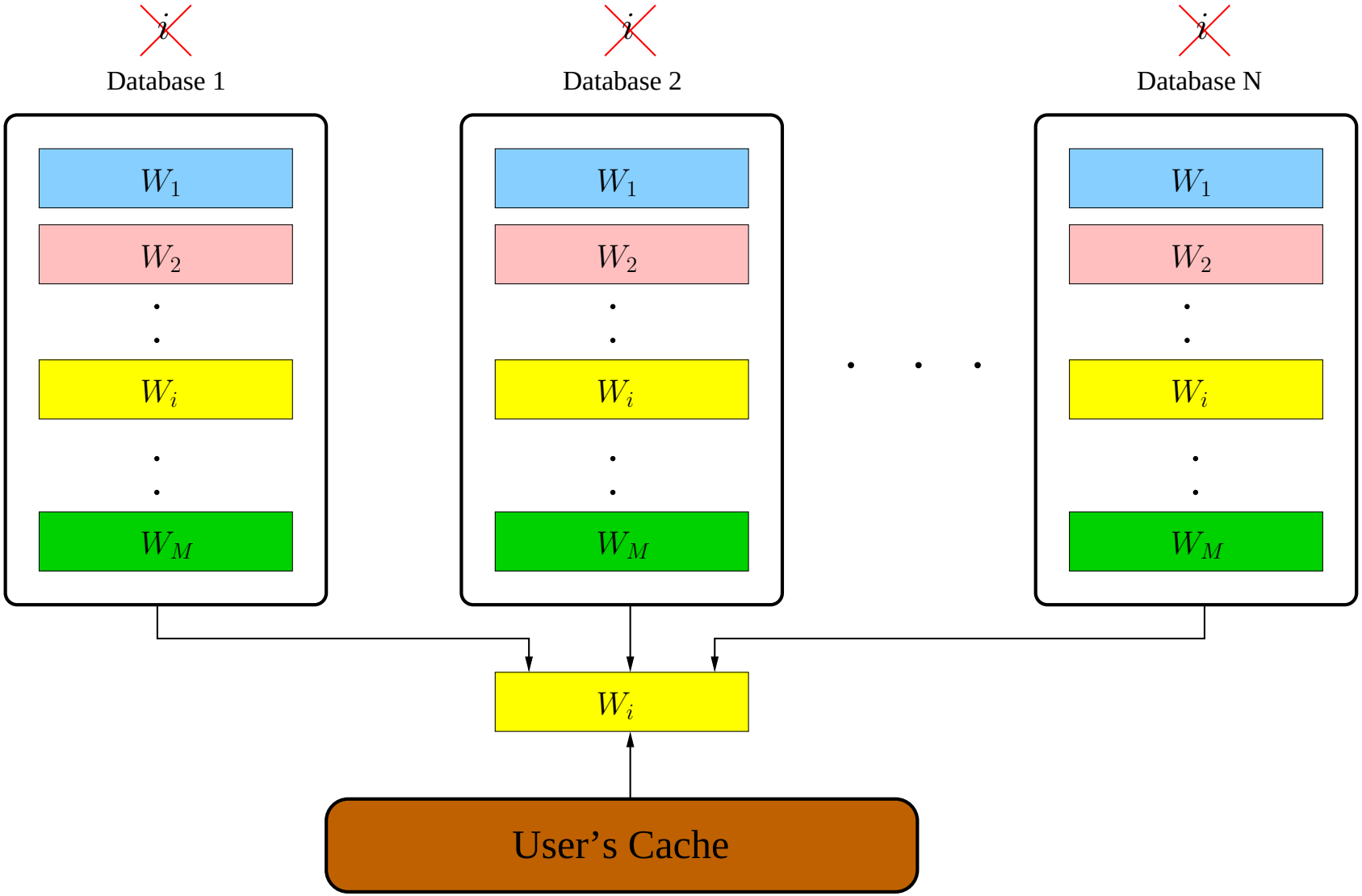
Database 1	Database 2
a_1, b_1, c_1	a_2, b_2, c_2
$a_3 + b_2$ $a_4 + c_2$ $b_3 + c_3$	$a_5 + b_1$ $a_6 + c_1$ $b_4 + c_4$
$a_7 + b_4 + c_4$	$a_8 + b_3 + c_3$
a_9, b_5, c_5	$a_{10} + b_5$ $a_{11} + c_5$ $b_6 + c_6$
$a_{12} + b_6 + c_6$	



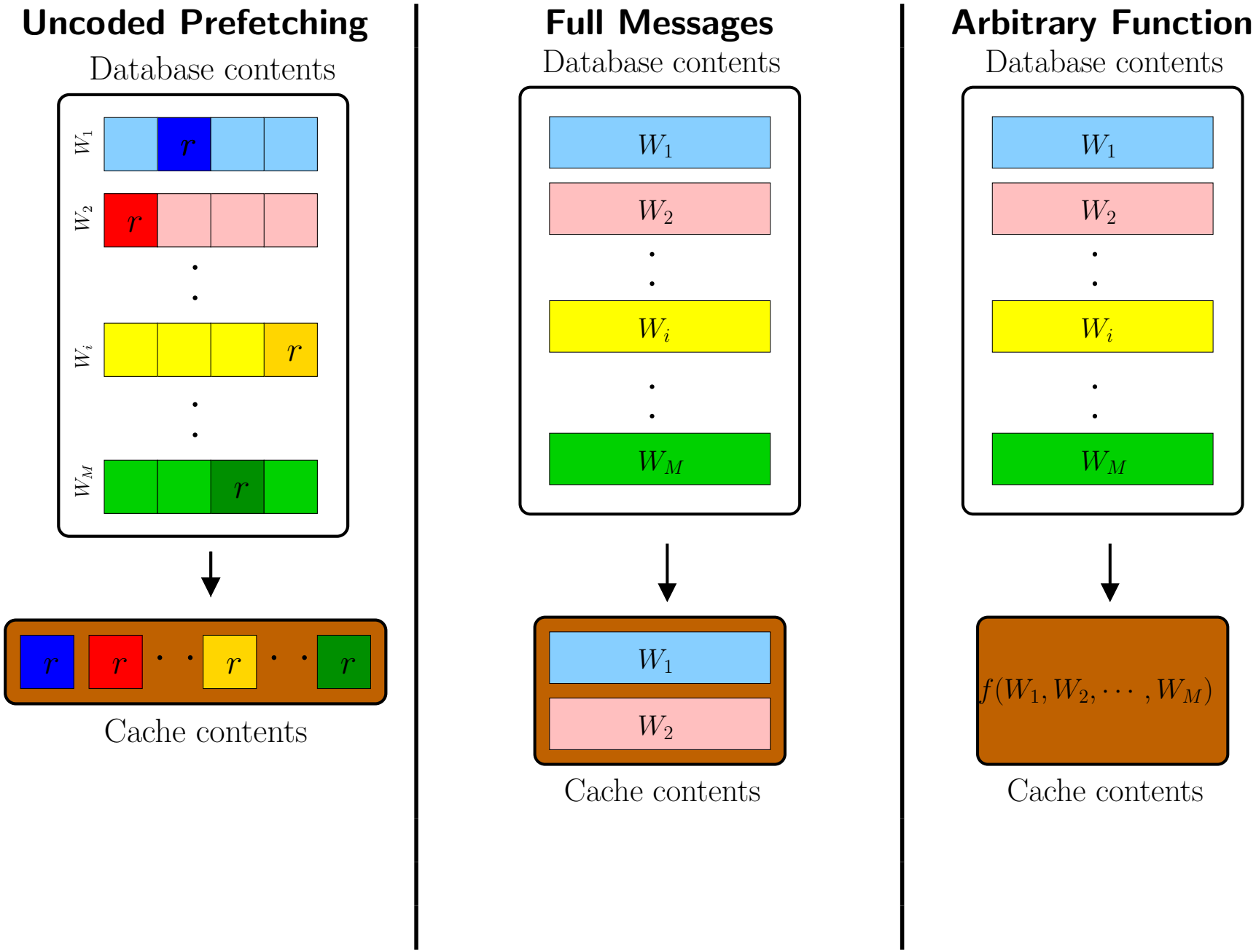
Non-Tight Example: $M = 4$, $N = 2$



Cache-Aided PIR



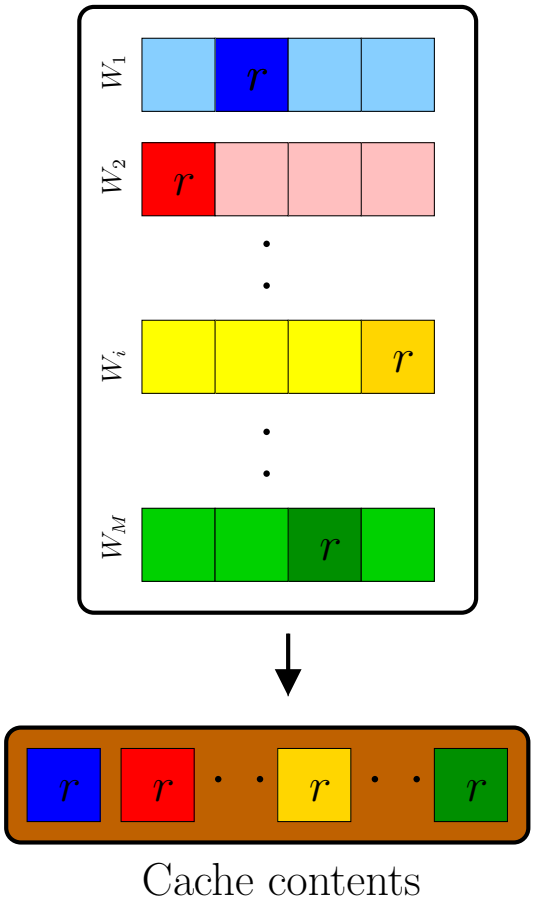
Variations of Cache-Aided PIR: Cache Format



Variations of Cache-Aided PIR: Cache Format

Uncoded Prefetching

Database contents

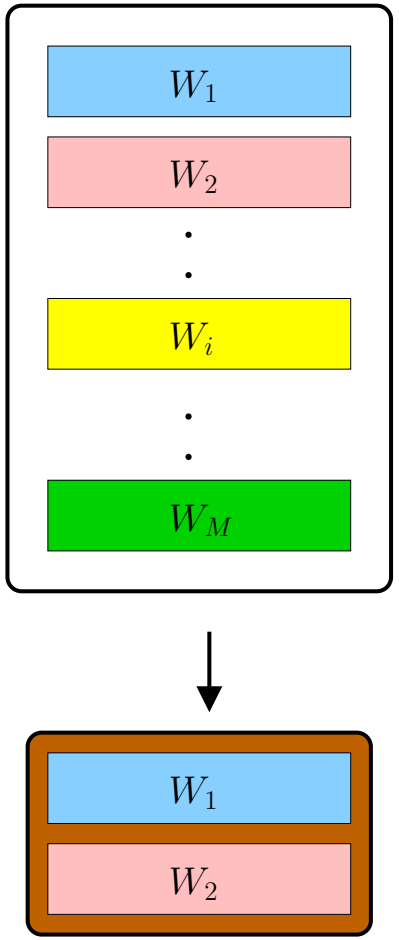


Cache contents

Wei et al., ISIT 2018

Full Messages

Database contents

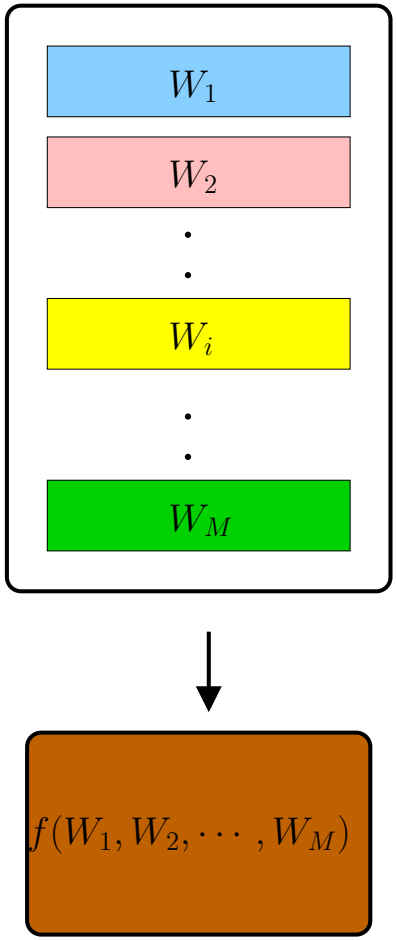


Cache contents

Kadhe et al., Allerton 2017
Chen et al., Arxiv 2017
Wei et al., CISS 2018

Arbitrary Function

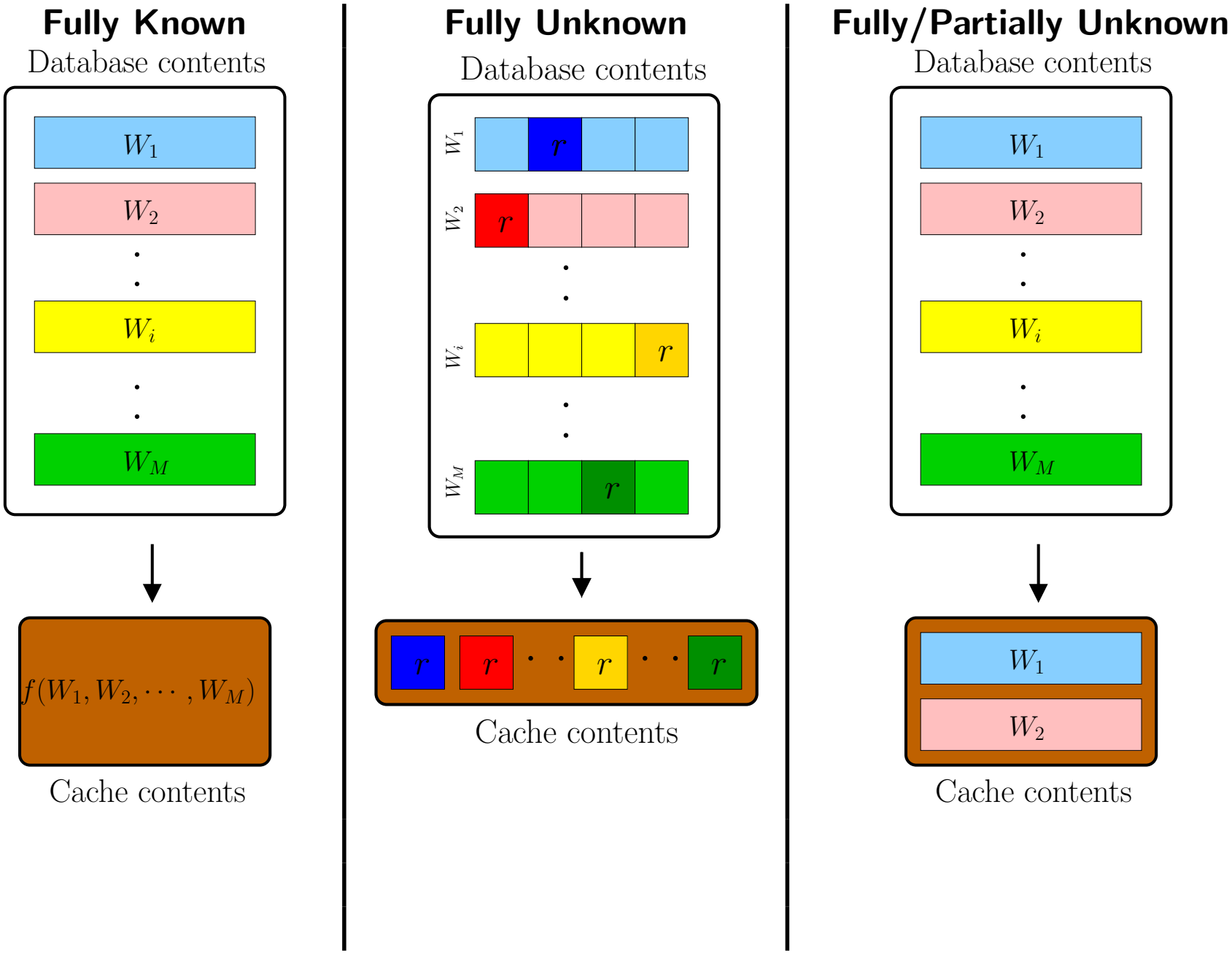
Database contents



Cache contents

Tandon, Allerton 2017

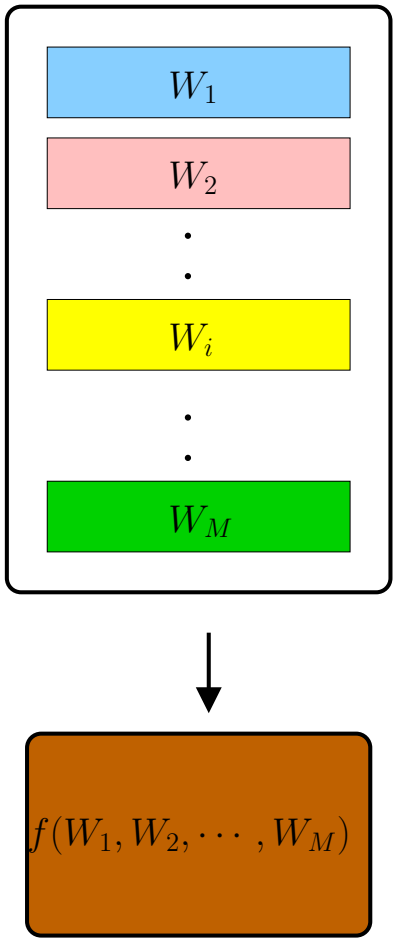
Variations of Cache-Aided PIR: Awareness of the Side Information



Variations of Cache-Aided PIR: Awareness of the Side Information

Fully Known

Database contents

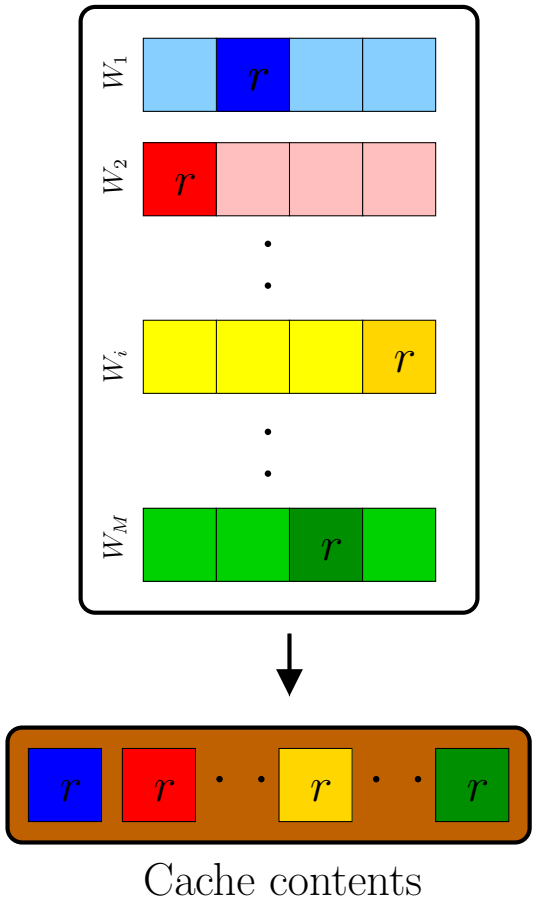


Cache contents

Tandon, Allerton 2017

Fully Unknown

Database contents

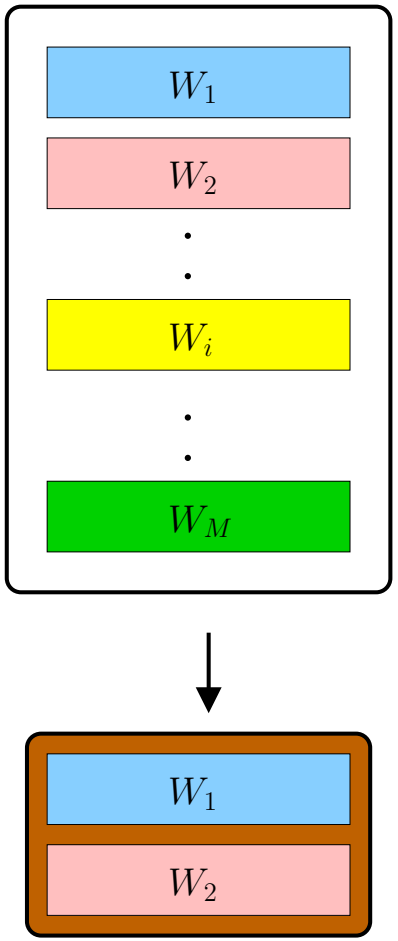


Cache contents

Wei et al., ISIT 2018

Fully/Partially Unknown

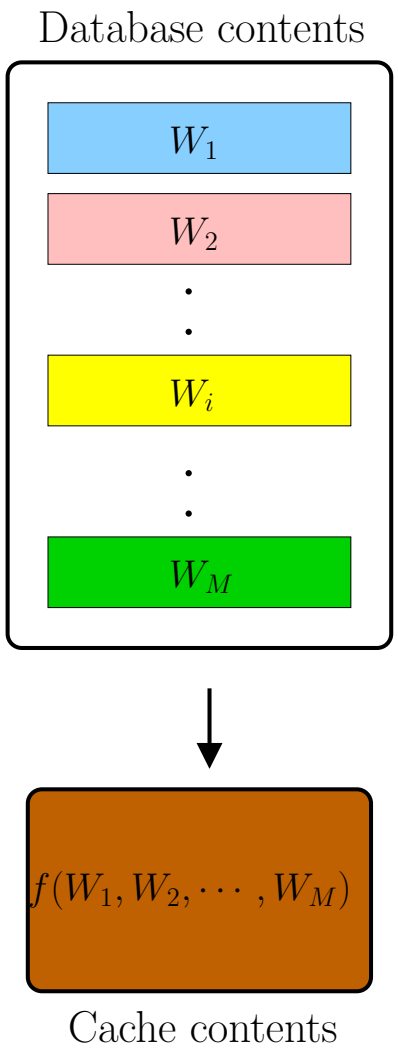
Database contents



Cache contents

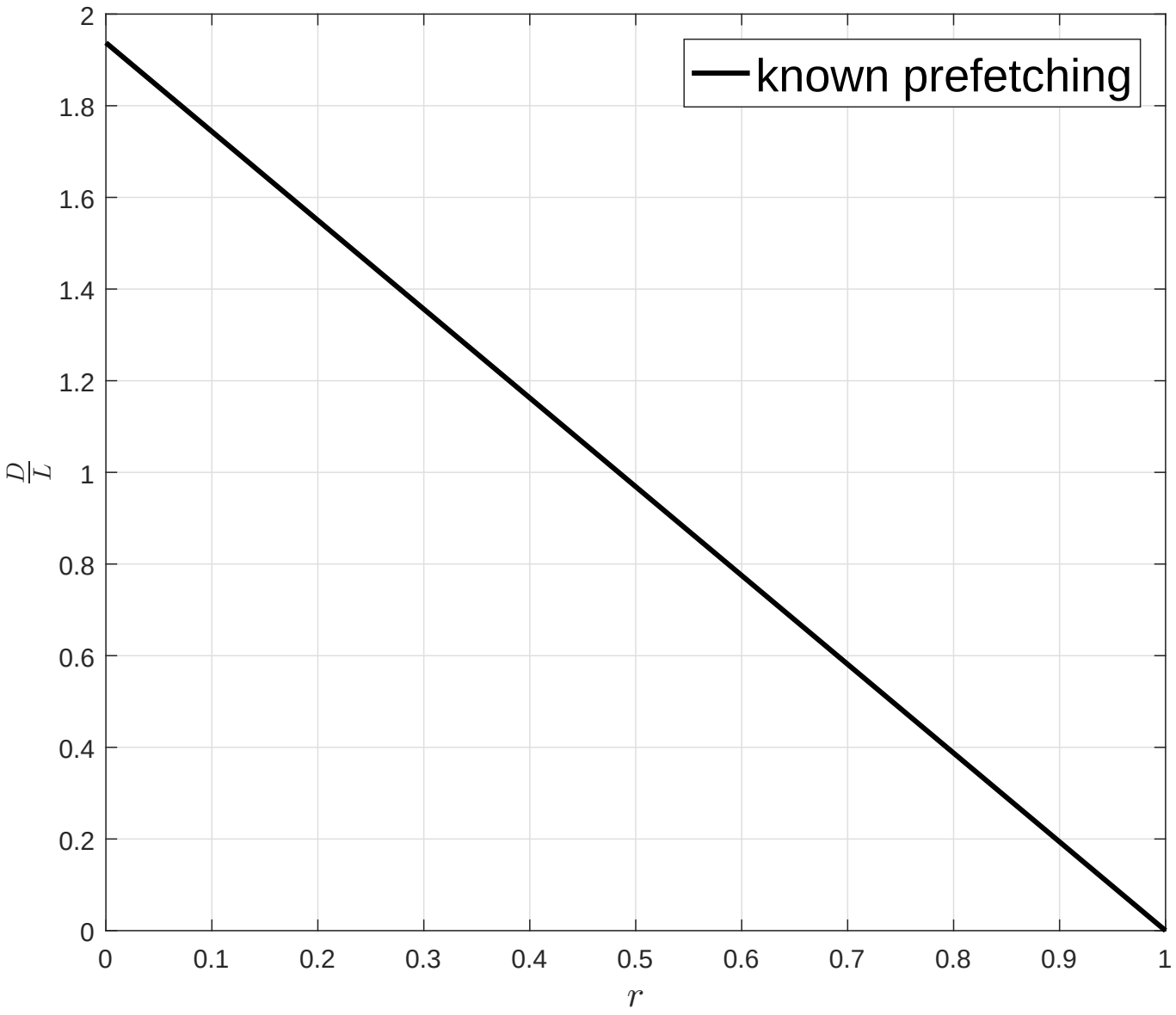
Kadhe et al., Allerton 2017
Chen et al., Arxiv 2017
Wei et al., CISS 2018

Arbitrary Function, Fully Known, Non-Private SI [Tandon][11]

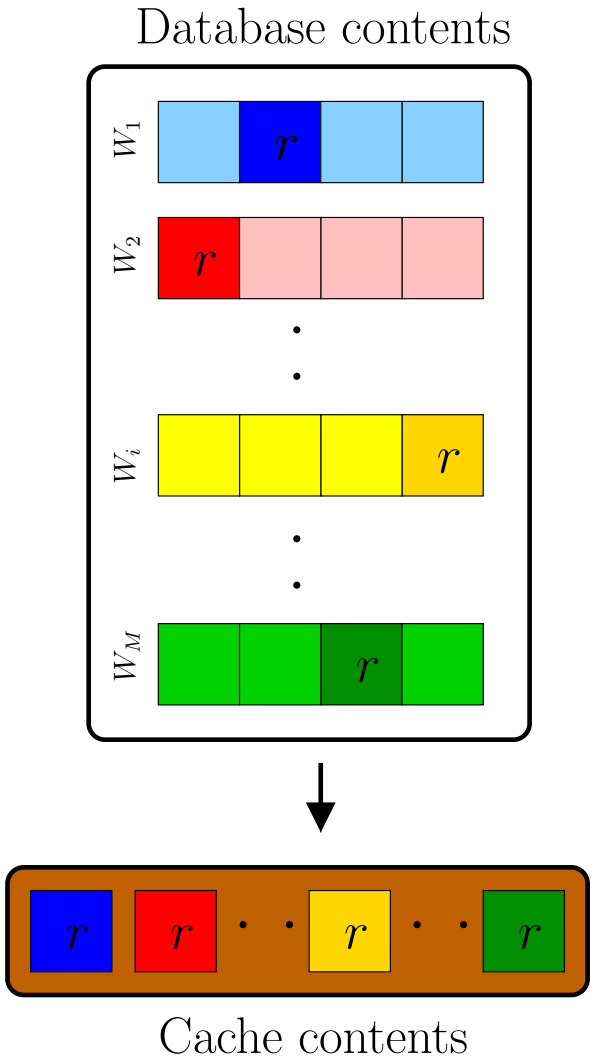


[11] R. Tandon. The capacity of cache aided private information retrieval. 2017. Available at arXiv: 1706.07035.

Memory-Download Cost Tradeoff: $M = 5$, $N = 2$



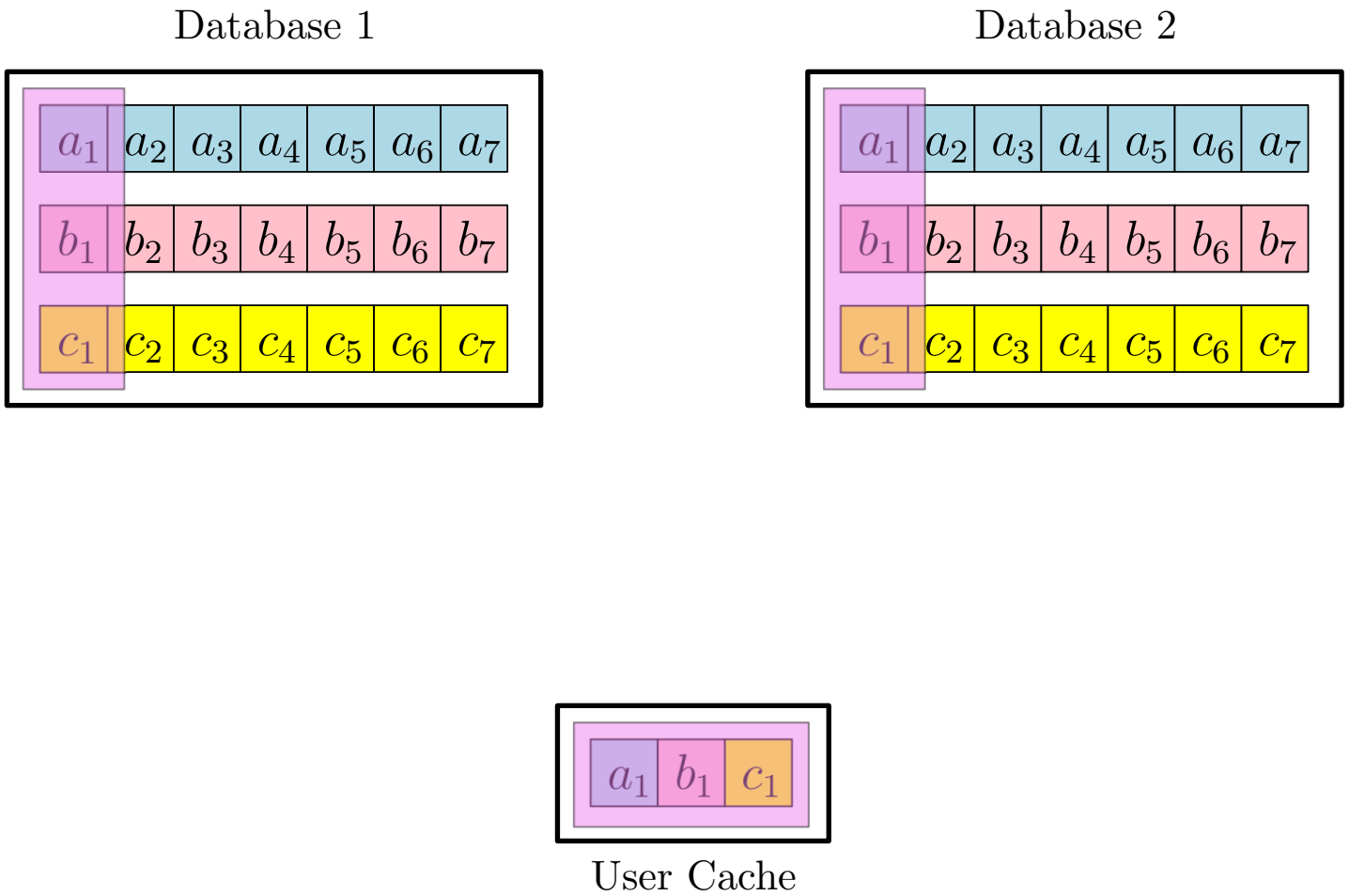
Uncoded, Fully Unknown, Non-Private SI [Wei-Banawan-Ulukus][12]



[12] Y.-P. Wei, K. Banawan, and S. Ulukus. Fundamental limits of cache-aided private information retrieval with unknown and uncoded prefetching. 2017. Available at arXiv:1709.01056.

Uncoded, Fully Unknown, Non-Private SI [Wei-Banawan-Ulukus][12]

- For each message, Lr out L bits are cached by the user.
- DBs do not know the cached bit indices.
- Ex: $N = 2$, $K = 3$, $r = \frac{1}{7}$.



[12] Y.-P. Wei, K. Banawan, and S. Ulukus. Fundamental limits of cache-aided private information retrieval with unknown and uncoded prefetching. 2017. Available at arXiv:1709.01056.

Converse Proof: New Tools and Lemmas

- **Lemma: Interference lower bound lemma**

$$D(r) - L(1-r) + o(L) \geq I\left(W_{k:M}; Q_{1:N}^{[k-1]}, A_{1:N}^{[k-1]} | W_{1:k-1}, Z, \mathbb{H}\right),$$

for $k \in \{2, \ldots, M\}$.

- **Lemma: Induction lemma**

$$I\left(W_{k:M}; Q_{1:N}^{[k-1]}, A_{1:N}^{[k-1]} | W_{1:k-1}, Z, \mathbb{H}\right)$$
$$\geq \frac{1}{N} I\left(W_{k+1:M}; Q_{1:N}^{[k]}, A_{1:N}^{[k]} | W_{1:k}, Z, \mathbb{H}\right) + \frac{L(1-r)}{N} - (M-k+1)Lr - o(L).$$

Converse Example: $M = 3$, $N = 2$

- $k = 2$ for interference lower bound, and apply induction lemma twice.

$$\begin{aligned} D(r) - L(1-r) &\geq I\left(W_{2:3}; Q_{1:2}^{[1]}, A_{1:2}^{[1]} | W_{1:1}, Z, \mathbb{H}\right) \\ &\geq \frac{1}{2} I\left(W_{3:3}; Q_{1:2}^{[2]}, A_{1:2}^{[2]} | W_{1:2}, Z, \mathbb{H}\right) + \frac{L(1-r)}{2} - 2Lr - o(L) \\ &\geq \frac{1}{2}\left[\frac{L(1-r)}{2} - Lr\right] + \frac{L(1-r)}{2} - 2Lr - o(L) \end{aligned}$$

Therefore, $\frac{D(r)}{L} \geq \frac{7}{4} - \frac{17}{4} r$.

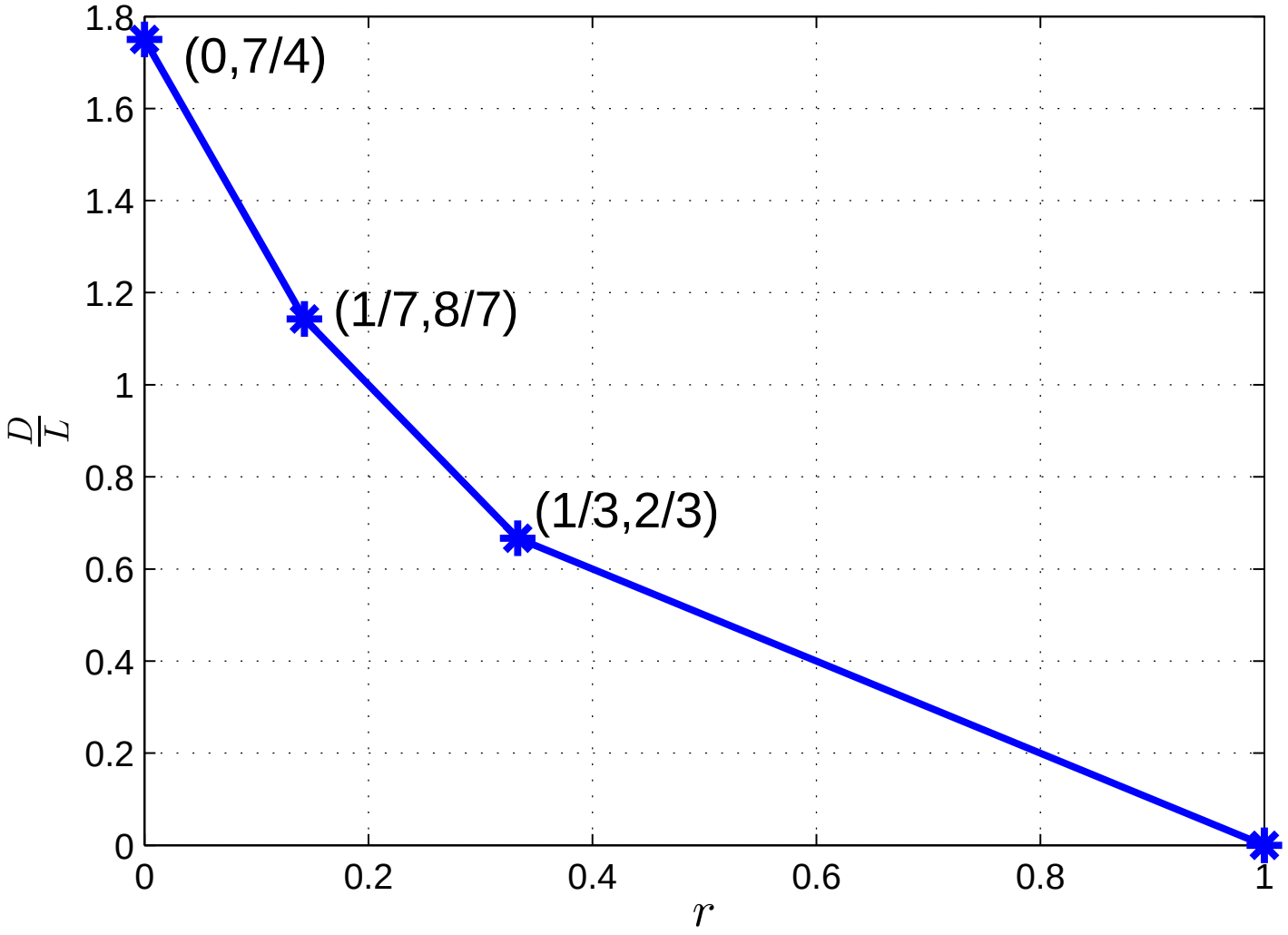
- $k = 3$ for interference lower bound, and apply induction lemma once.

$$D(r) - L(1-r) \geq I\left(W_{3:3}; Q_{1:2}^{[2]}, A_{1:2}^{[2]} | W_{1:2}, Z, \mathbb{H}\right) \geq \frac{L(1-r)}{2} - Lr - o(L)$$

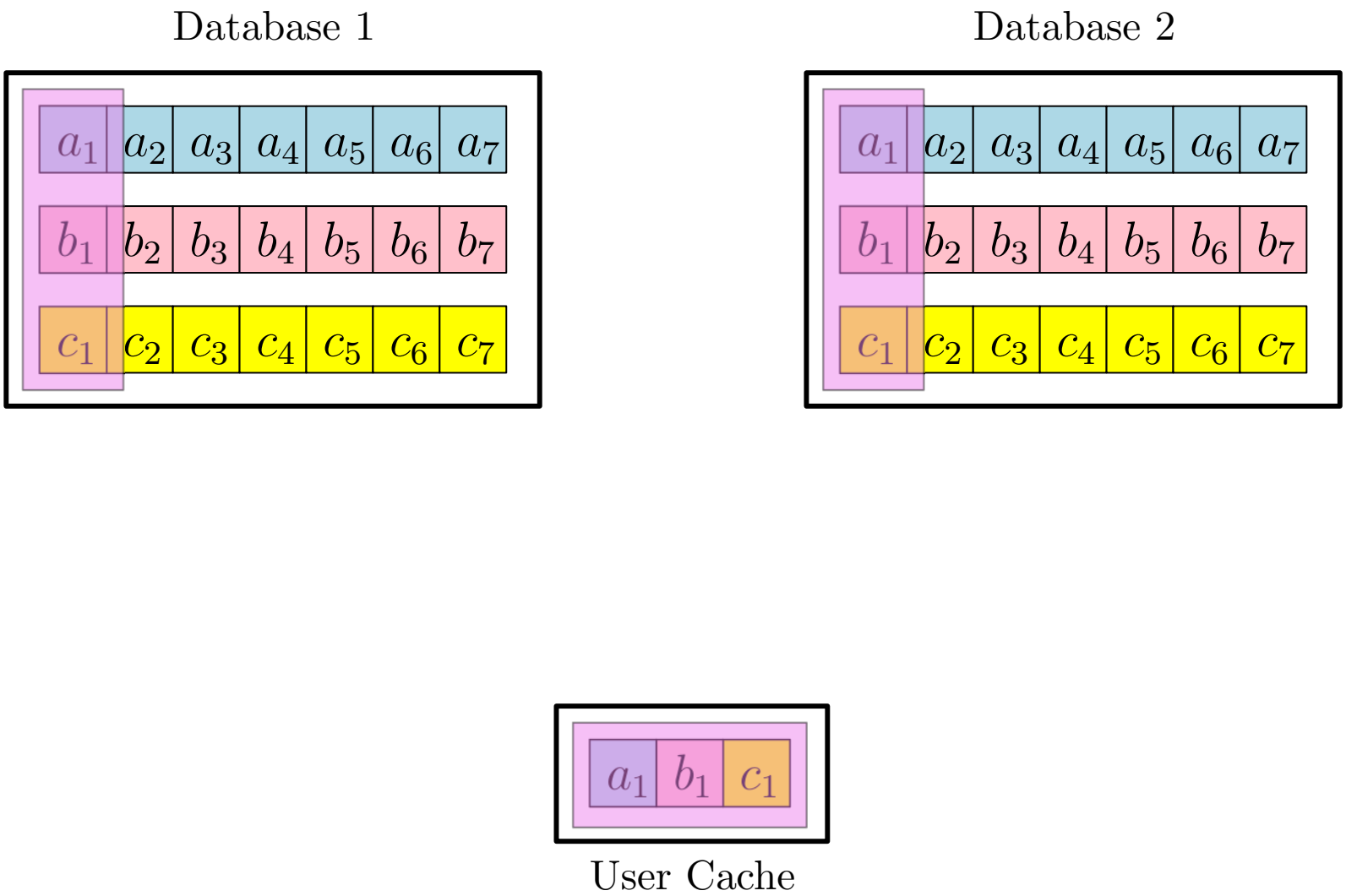
Therefore, $\frac{D(r)}{L} \geq \frac{3}{2} - \frac{5}{2} r$.

- Non-negativity of mutual information for interference lower bound.
Therefore, $\frac{D(r)}{L} \geq 1 - r$.

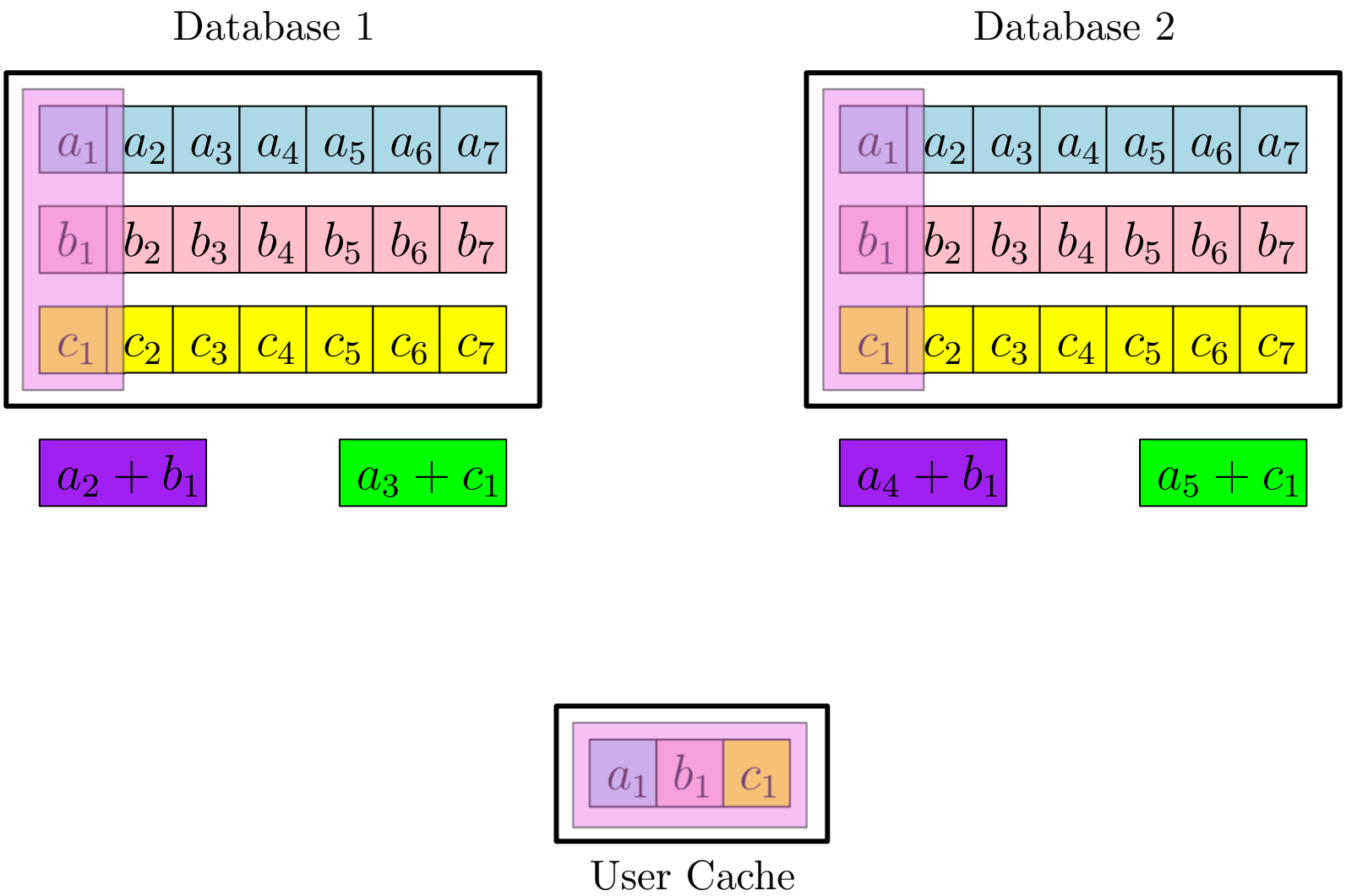
Optimal Normalized Download Cost for $M = 3$, $N = 2$



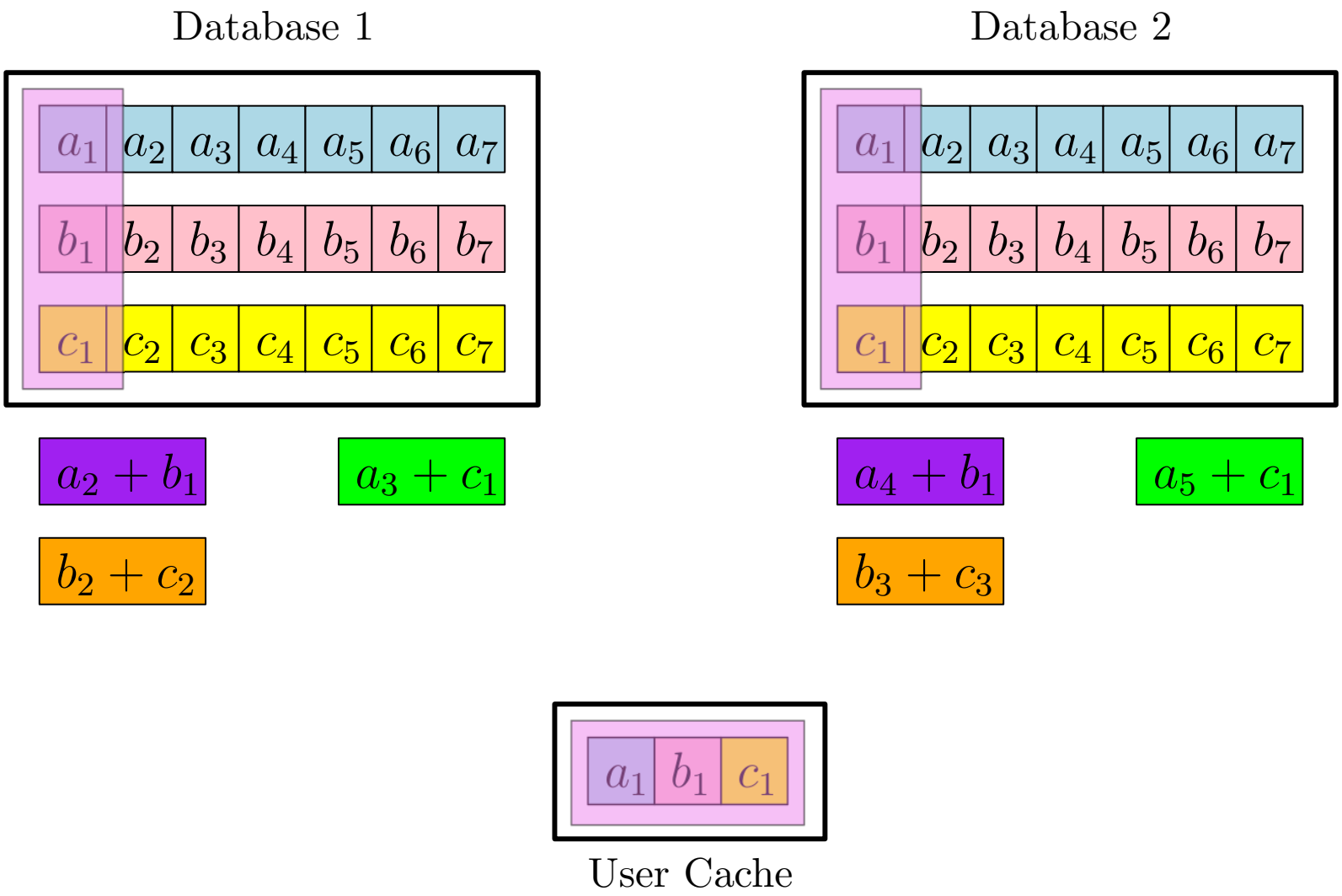
Achievability Example: $M = 3$, $N = 2$ and $r = \frac{1}{7}$



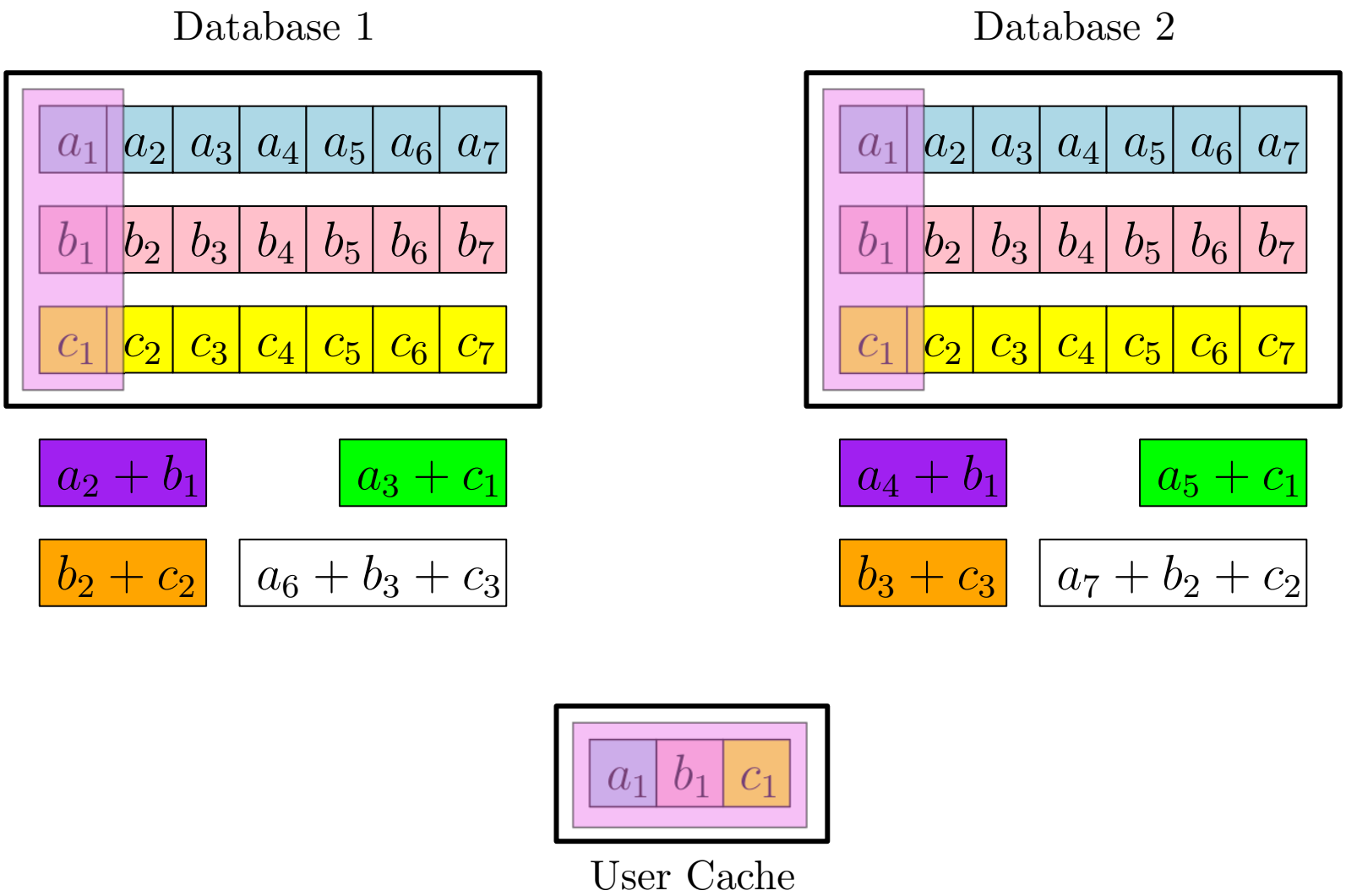
Example: $M = 3$, $N = 2$ and $r = \frac{1}{7}$: Queries



Example: $M = 3$, $N = 2$ and $r = \frac{1}{7}$: Queries



Example: $M = 3$, $N = 2$ and $r = \frac{1}{7}$: Queries



Example: $M = 3$, $N = 2$ and $r = \frac{1}{7}$: Query Table

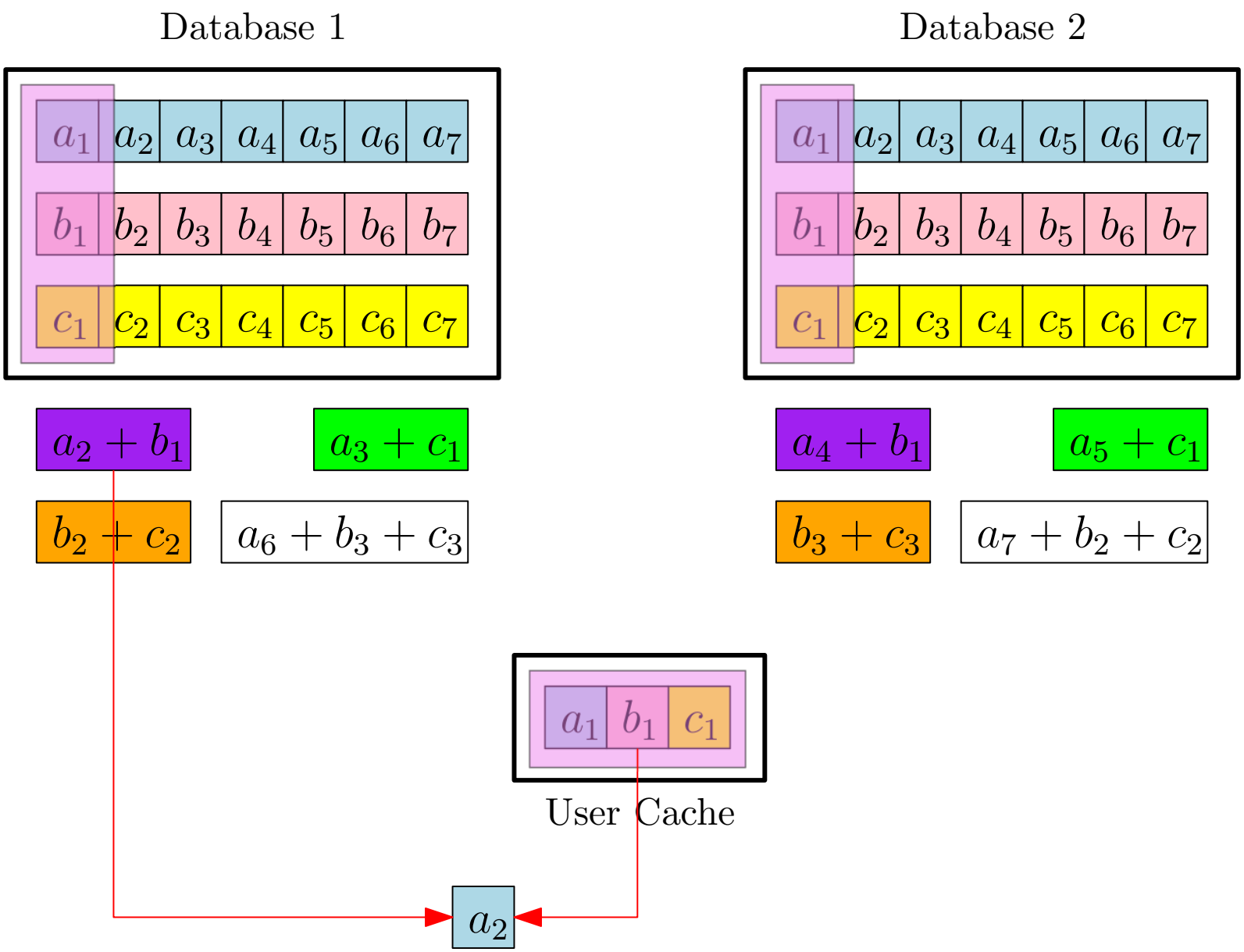
s	DB1	DB2
$s = 1$	$a_2 + b_1$	$a_4 + b_1$
	$a_3 + c_1$	$a_5 + c_1$
	$b_2 + c_2$	$b_3 + c_3$
	$a_6 + b_3 + c_3$	$a_7 + b_2 + c_2$

$$Z = (a_1, b_1, c_1)$$

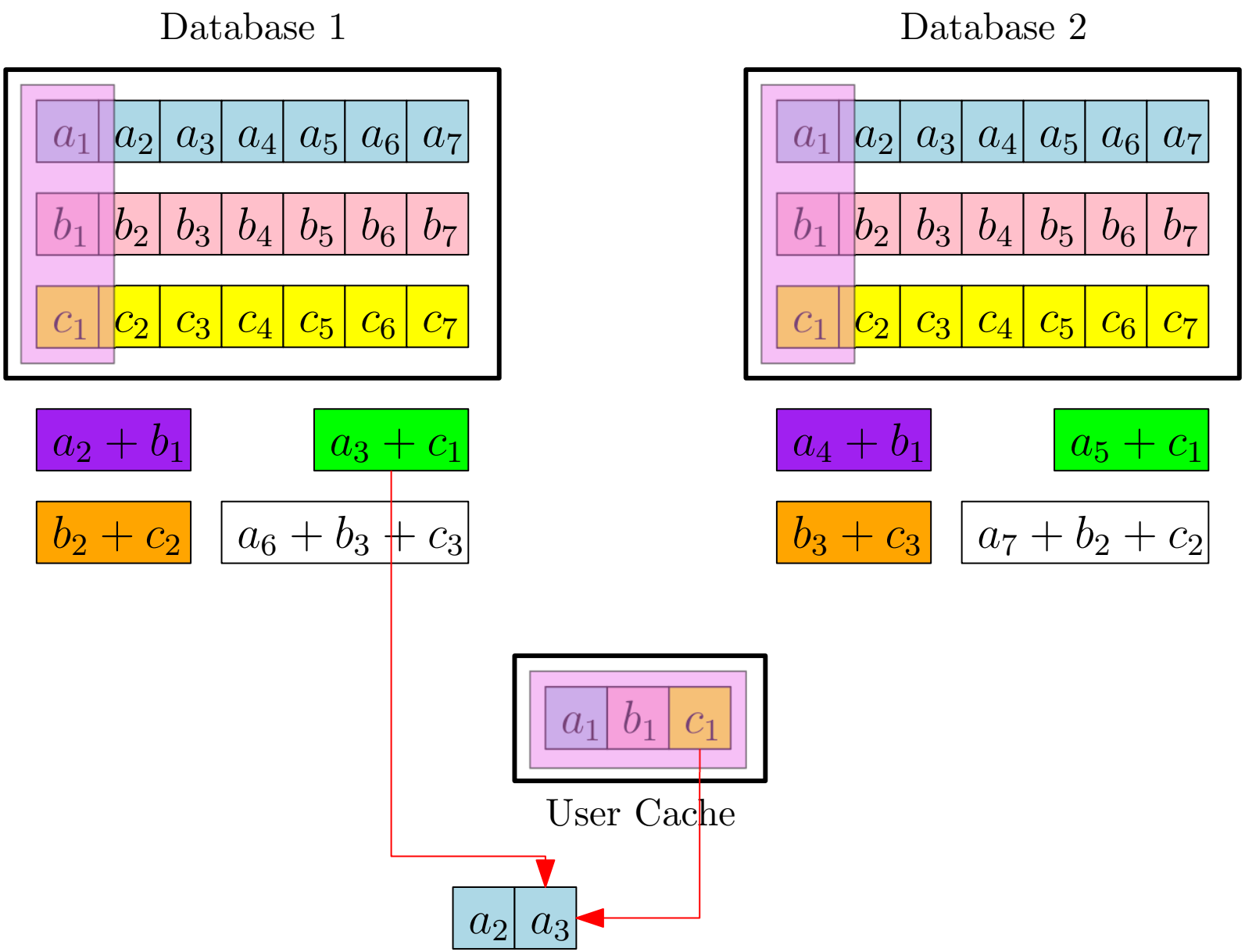
- Download cost:

$$\frac{D}{L} = \frac{8}{7}$$

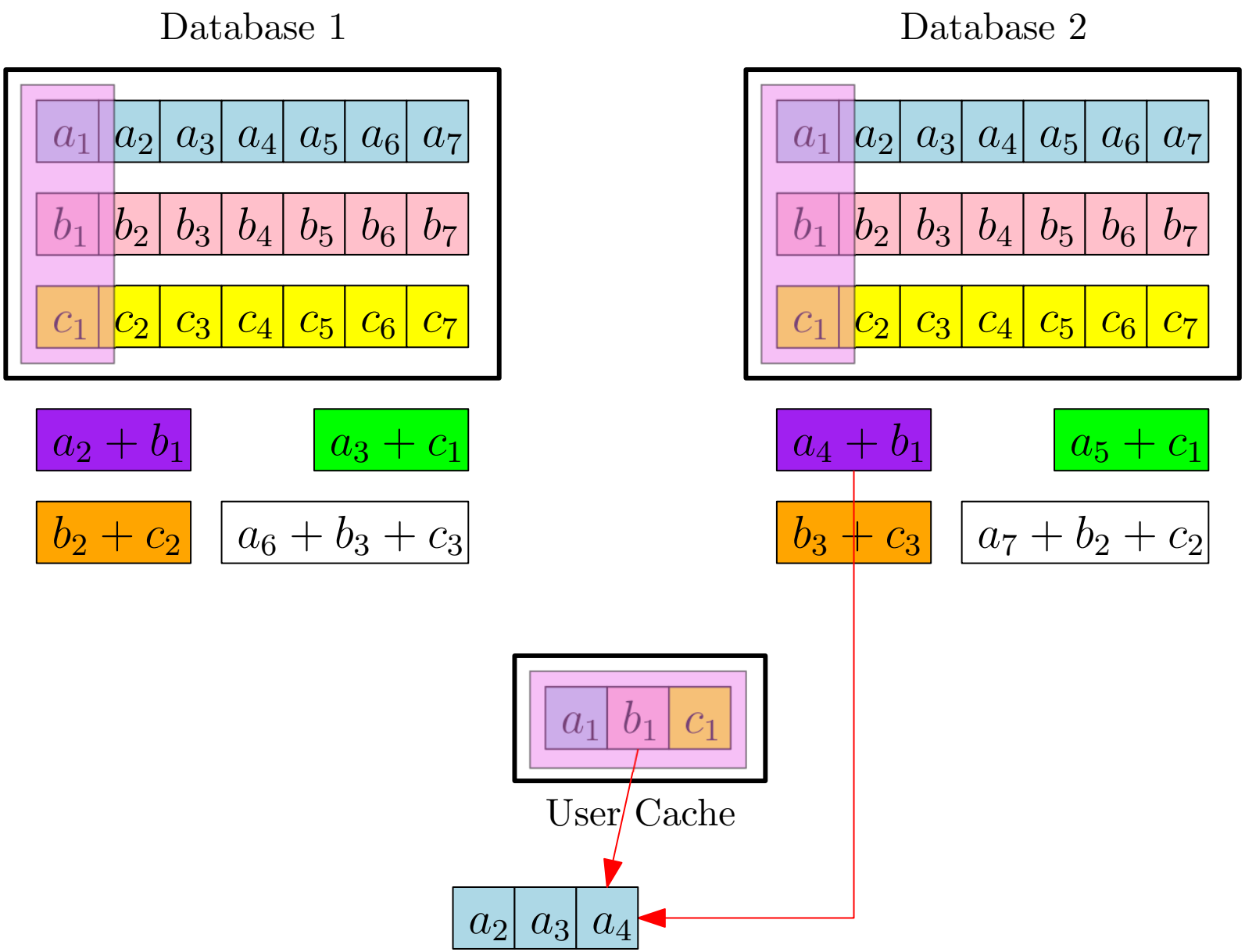
Example: $M = 3$, $N = 2$ and $r = \frac{1}{7}$: Decoding



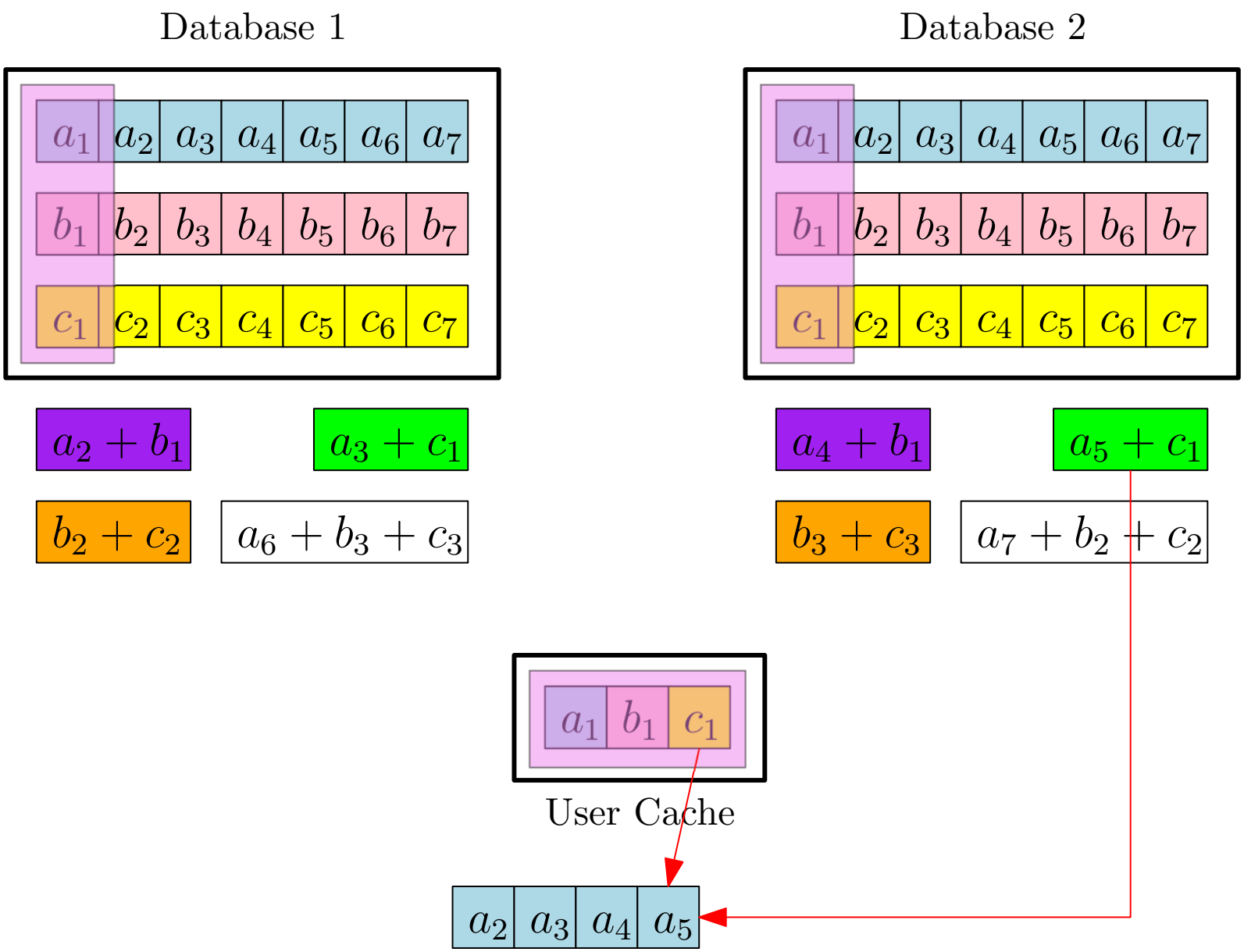
Example: $M = 3$, $N = 2$ and $r = \frac{1}{7}$: Decoding



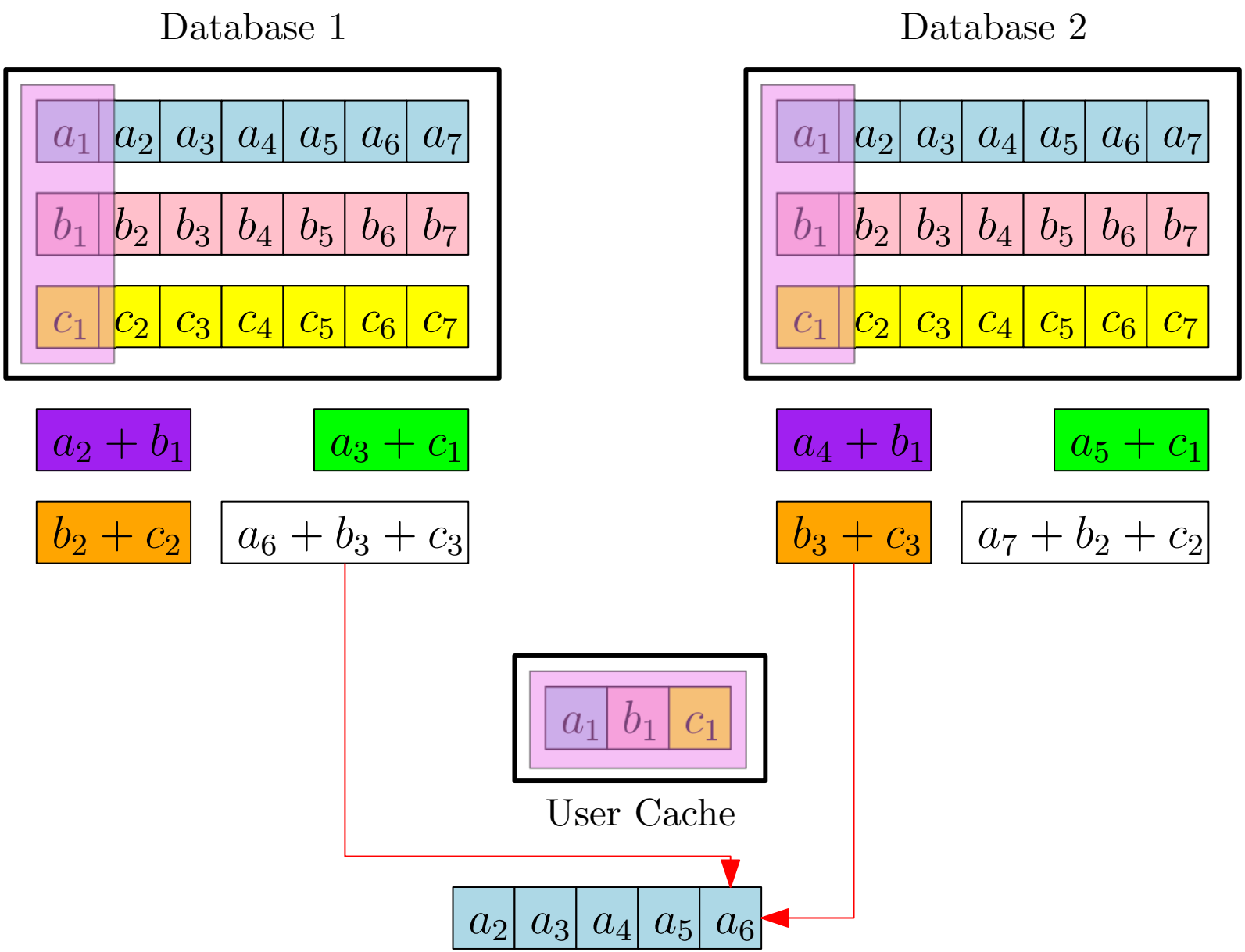
Example: $M = 3$, $N = 2$ and $r = \frac{1}{7}$: Decoding



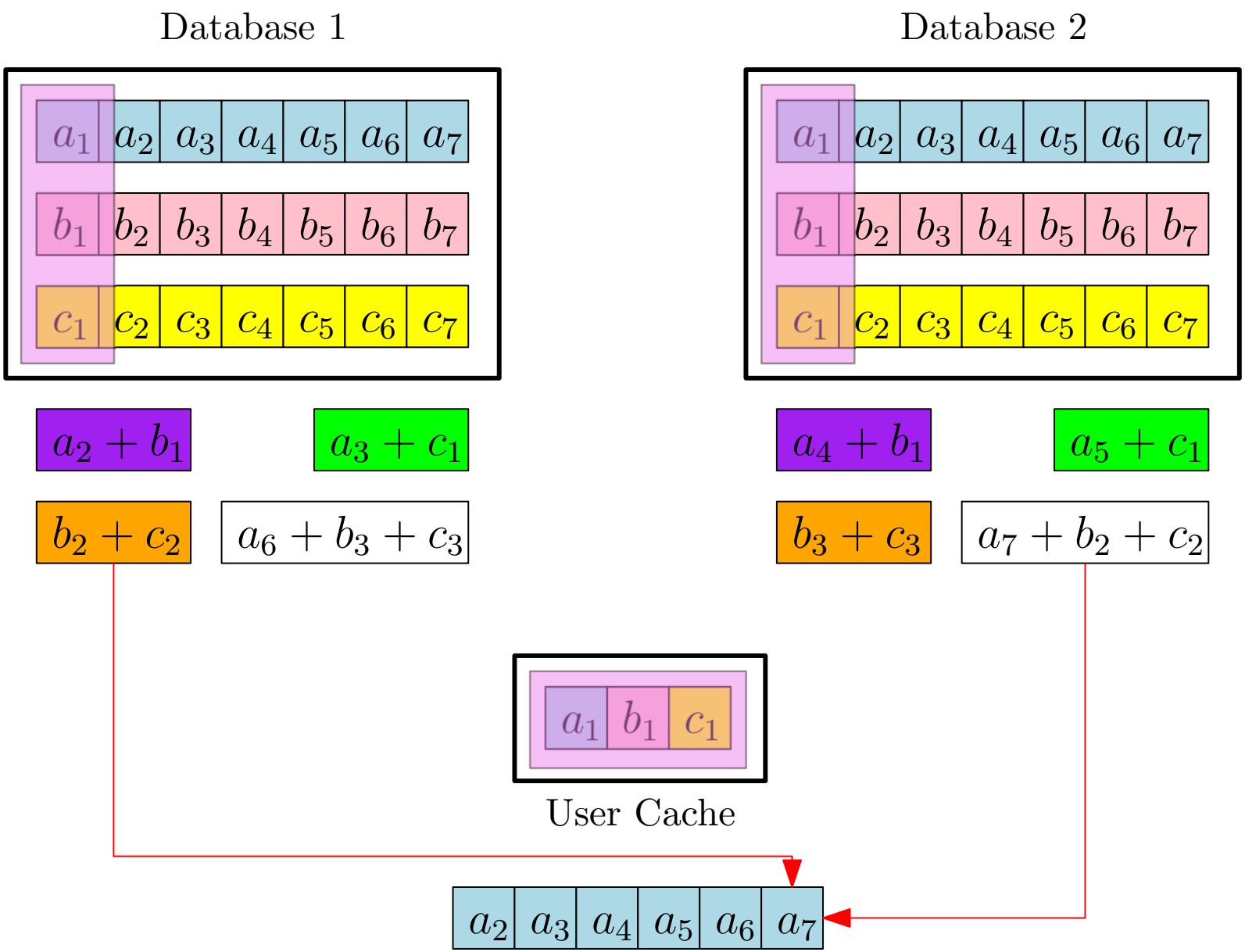
Example: $M = 3$, $N = 2$ and $r = \frac{1}{7}$: Decoding



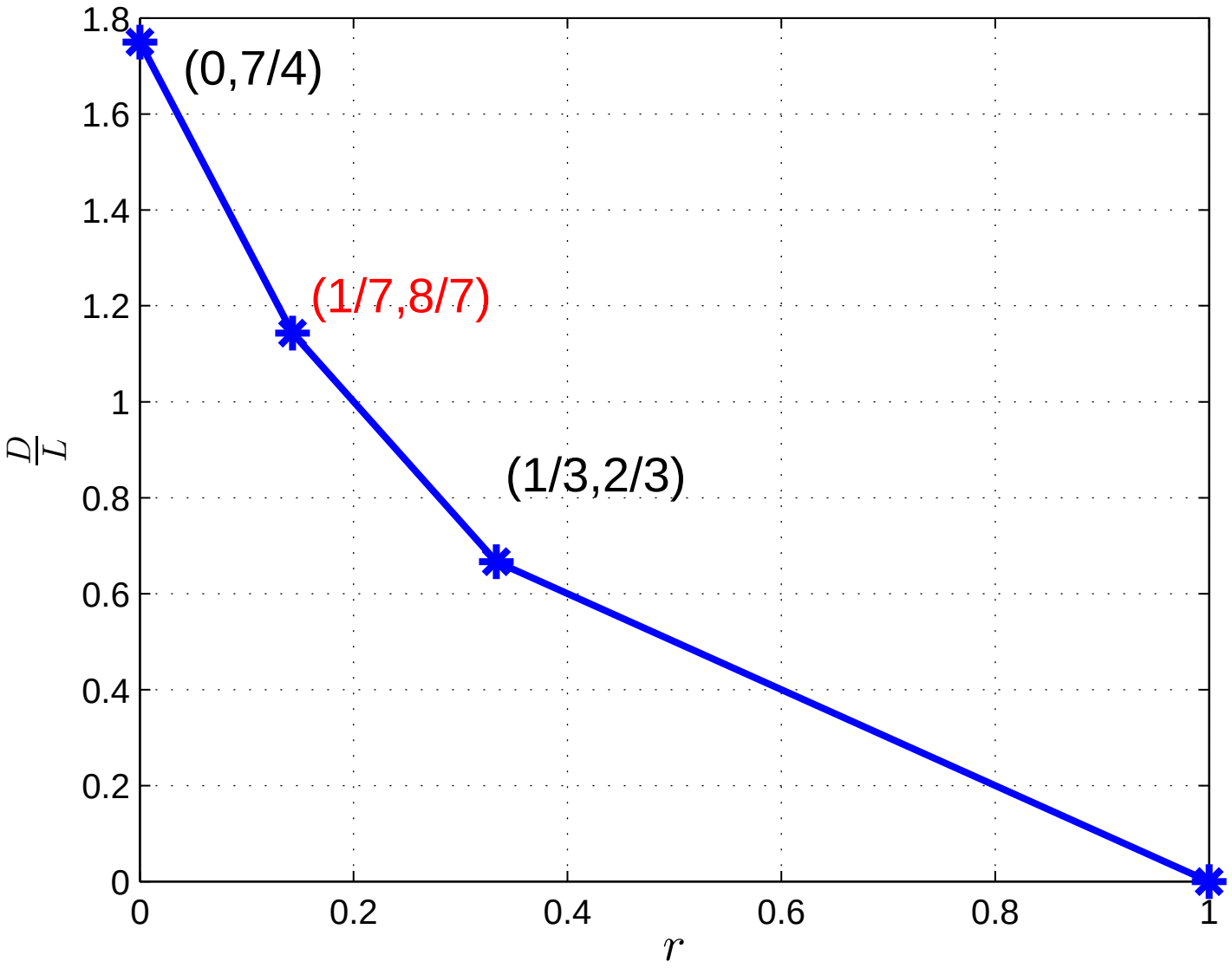
Example: $M = 3$, $N = 2$ and $r = \frac{1}{7}$: Decoding



Example: $M = 3$, $N = 2$ and $r = \frac{1}{7}$: Decoding



Optimal Normalized Download Cost for $M = 3$, $N = 2$



Example: $M = 3$, $N = 2$ and $r = \frac{1}{3}$: Query Table

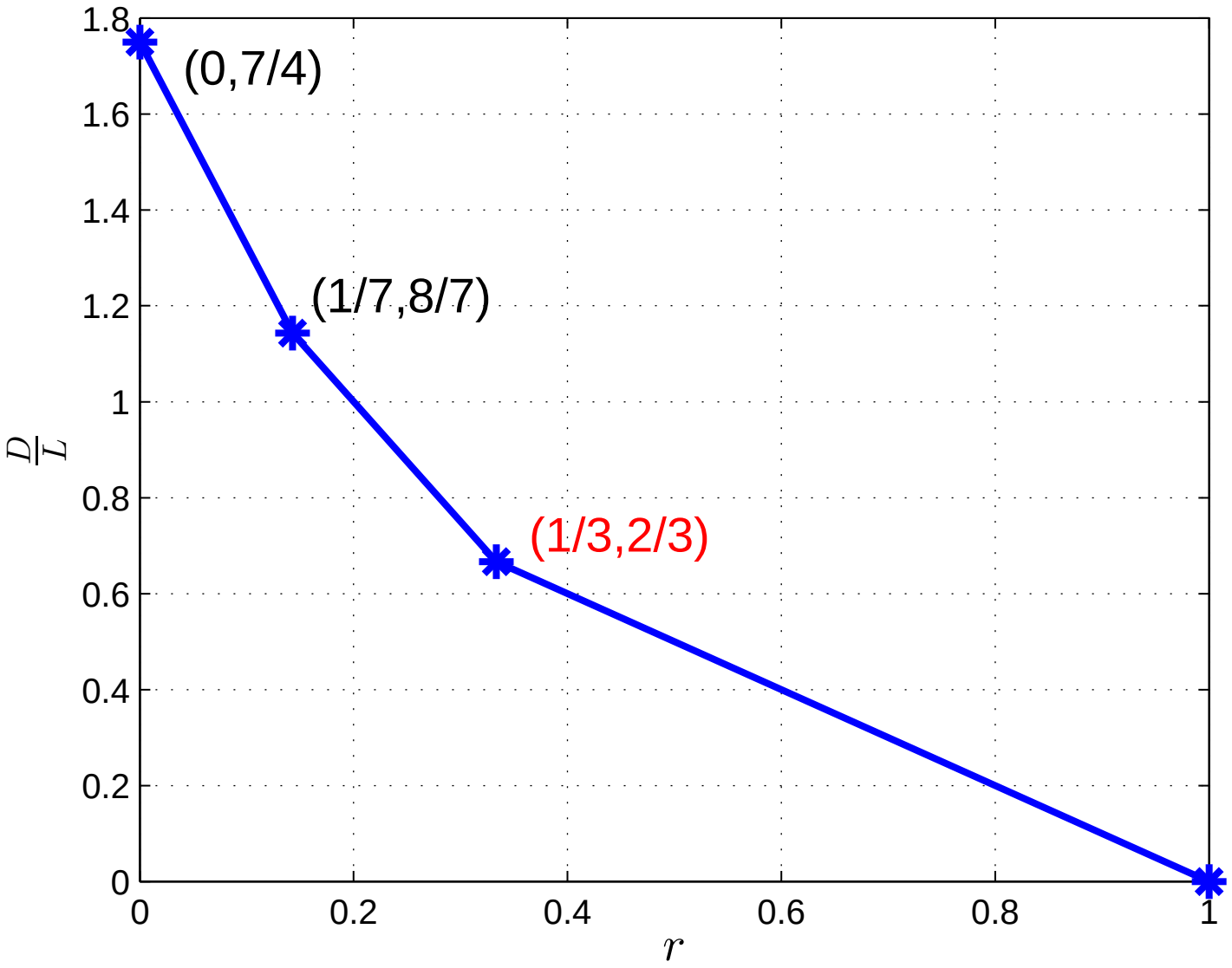
s	DB1	DB2
$s = 2$	$a_2 + b_1 + c_1$	$a_3 + b_1 + c_1$

$Z_1 = (a_1, b_1, c_1)$

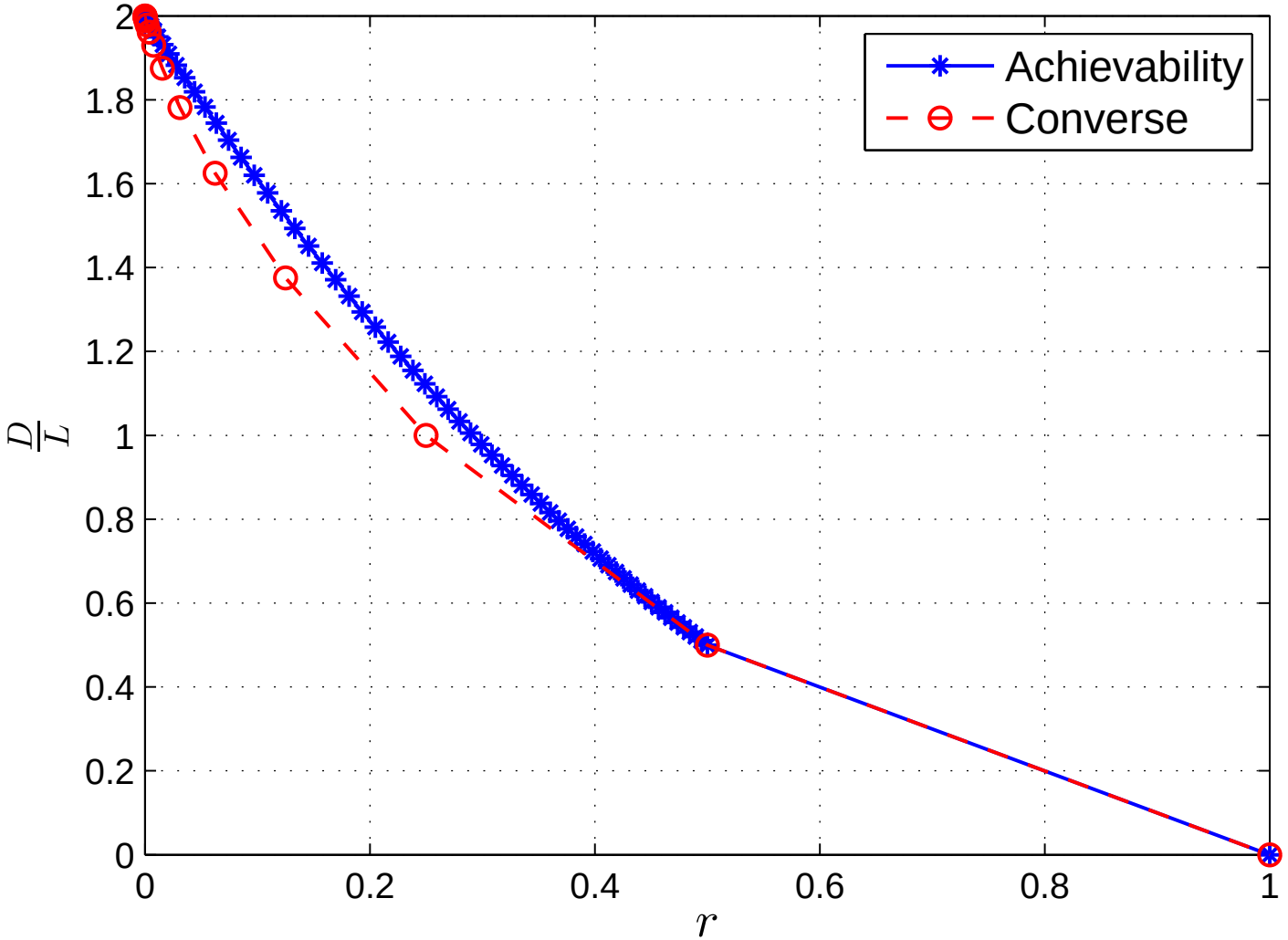
- Download cost:

$$\frac{D}{L} = \frac{2}{3}$$

Example: Optimal Normalized Download Cost for $M = 3$, $N = 2$

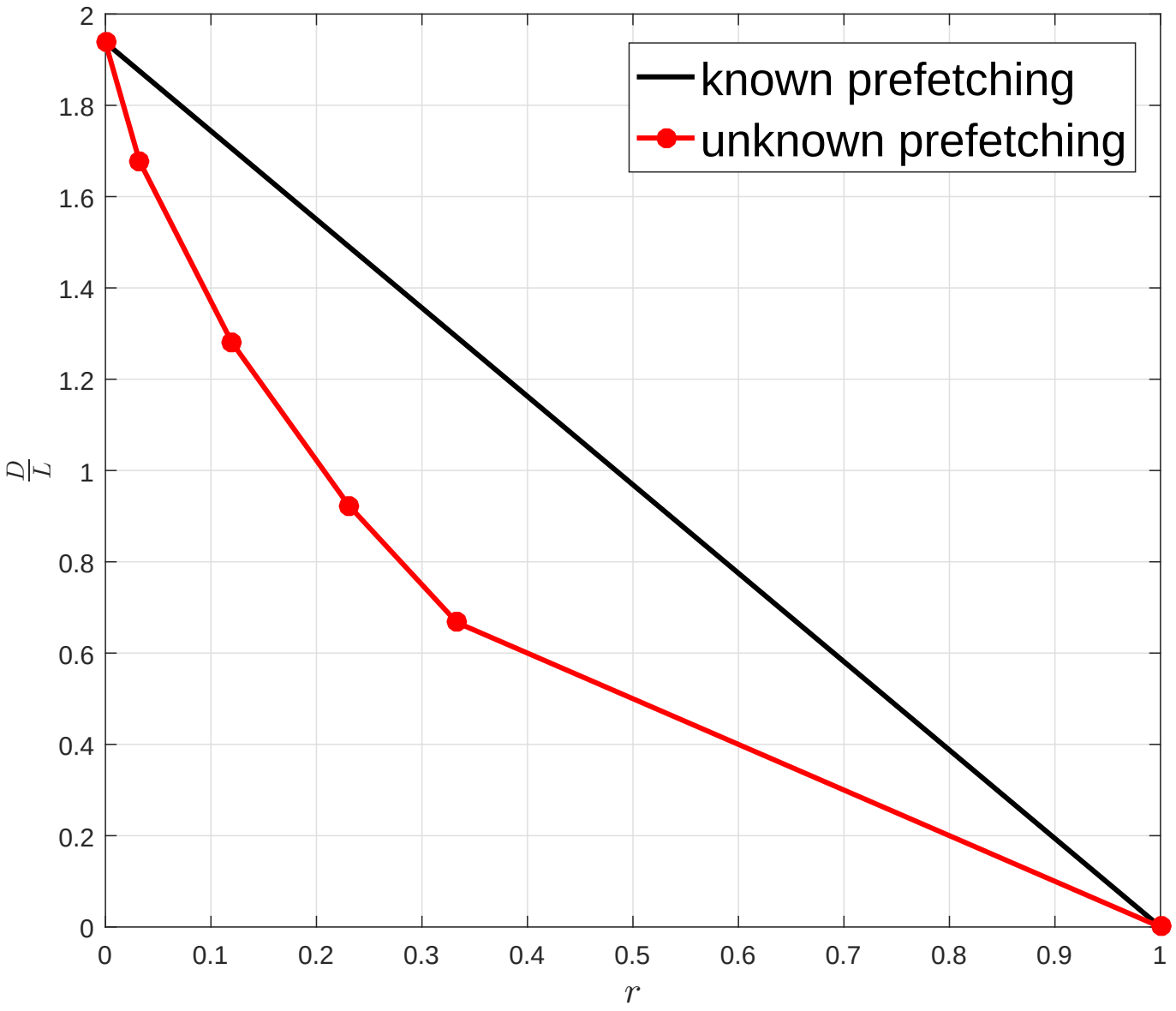


Normalized Download Cost for $M = 100$, $N = 2$



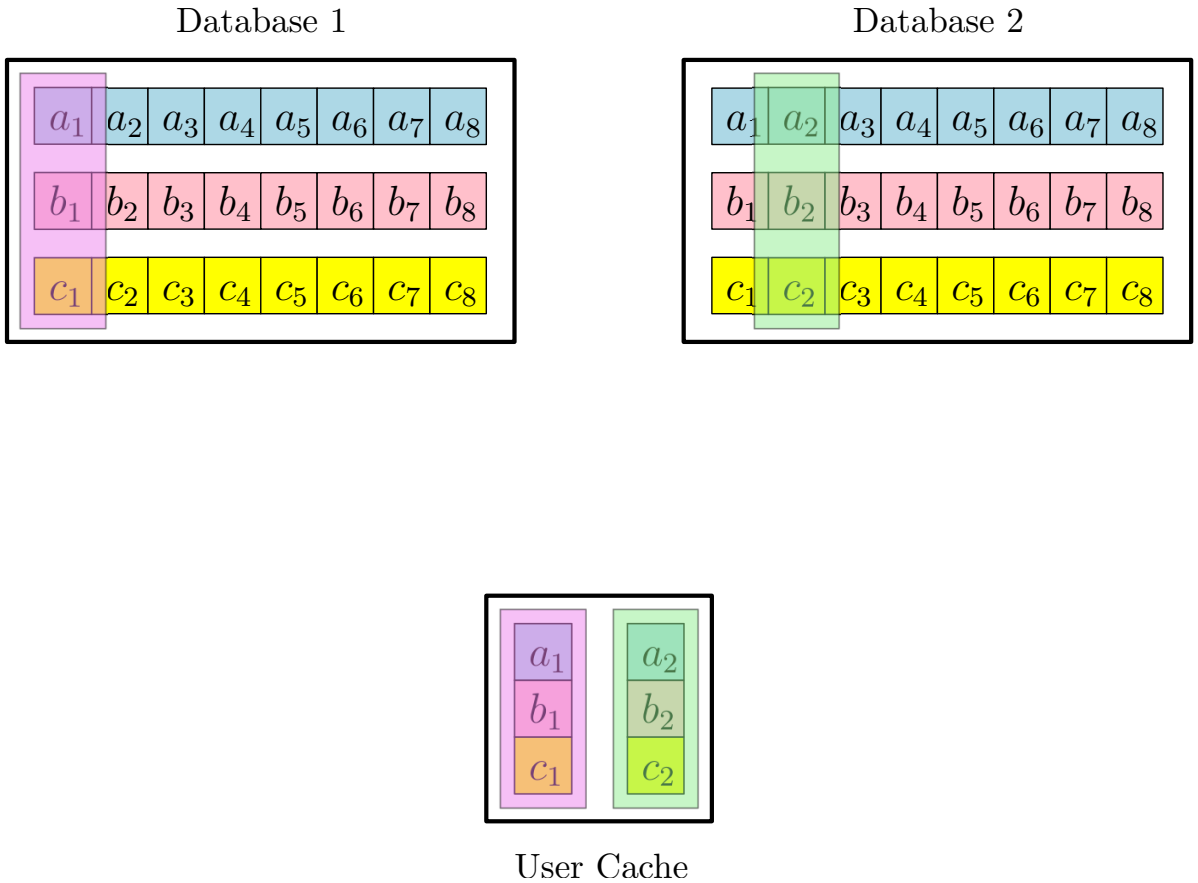
- **Worst additive gap is $\frac{1}{6}$.**

Awareness gain: $M = 5,\ N = 2$



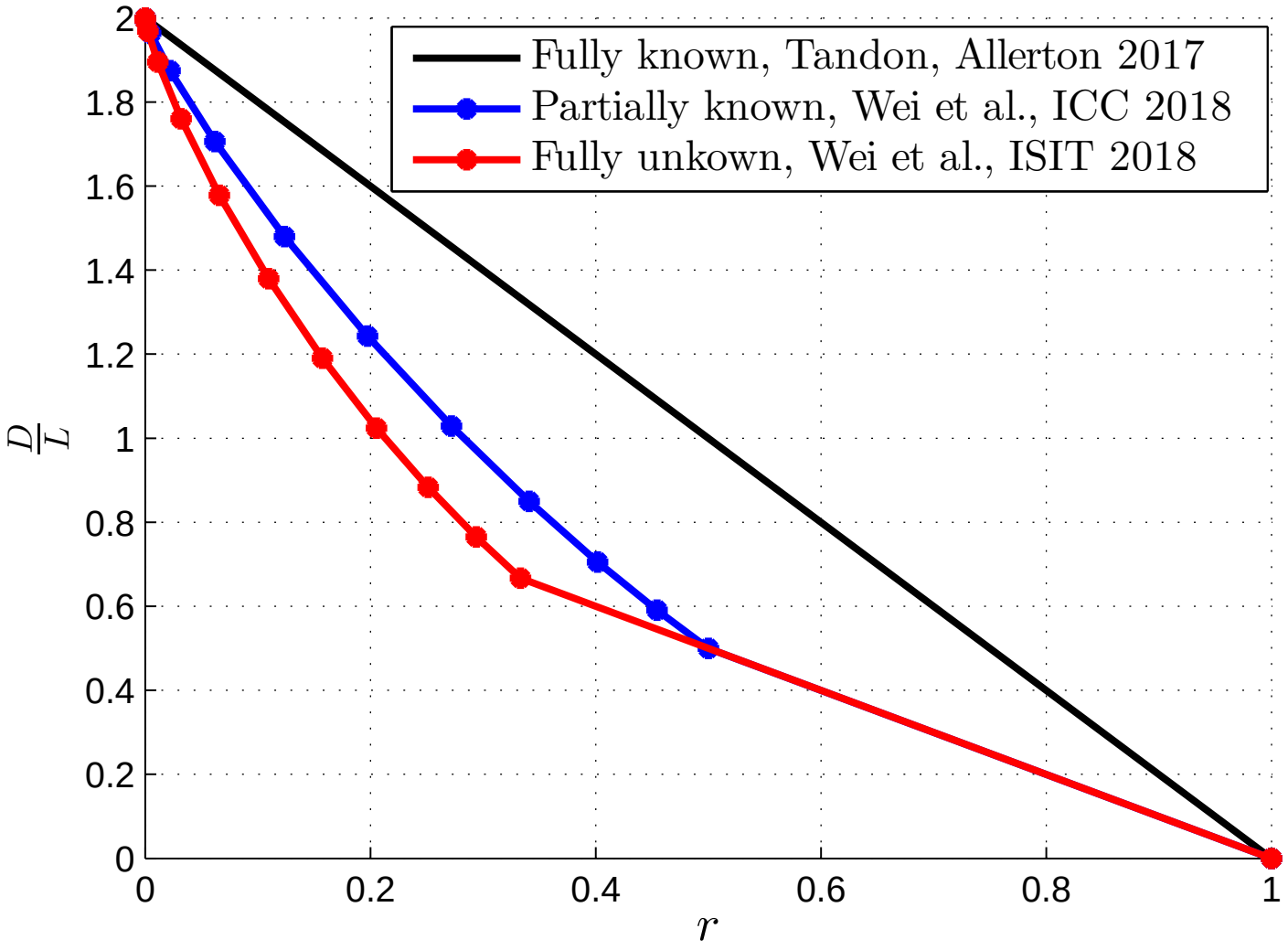
Partially Unknown Uncoded Prefetching [Wei-Banawan-Ulukus][13]

- For each message, Lr out L bits are cached by the user.
- Caches $\frac{MLr}{N}$ bits from DB1, $\frac{MLr}{N}$ bits from DB2, ..., $\frac{MLr}{N}$ bits from DBN.
- DBn knows the $\frac{MLr}{N}$ bits the user has cached from DBn.
- DBn does not know other bits the user cached from other databases.

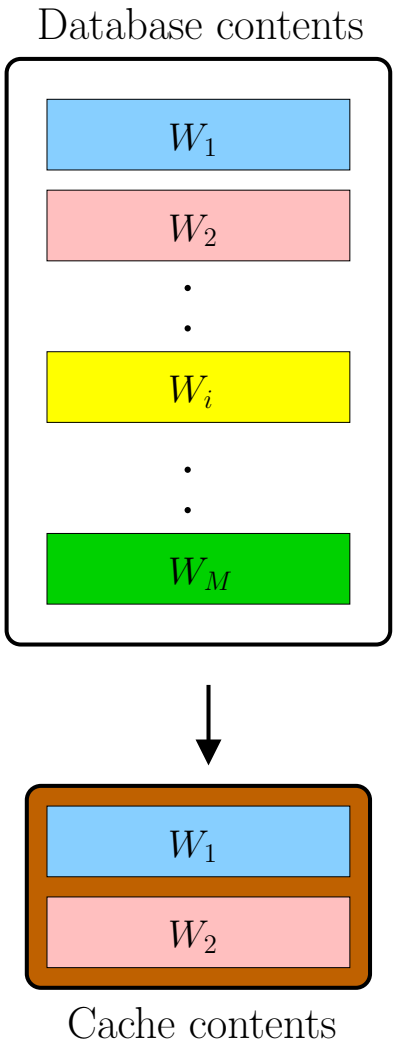


[13] Y.-P. Wei, K. Banawan and S. Ulukus, Cache-Aided Private Information Retrieval with Partially Known Uncoded Prefetching: Fundamental Limits, IEEE Jour. on Selected Areas in Communications, to appear. Available at arXiv: 1712.07021

Awareness Gain Comparison: $M = 12$, $N = 2$



Full Messages, Fully/Partially Unknown, Private SI [Chen-Wang-Jafar[14] & Wei-Banawan-Ulukus[15]]



[14] Z. Chen, Z. Wang, and S. Jafar. The capacity of private information retrieval with private side information. 2017. Available at arXiv:1709.03022.

[15] Y.-P. Wei, K. Banawan, and S. Ulukus. The capacity of private information retrieval with partially known private side information. 2017. Available at arXiv:1710.00809.

Full Messages, Fully/Partially Unknown, Private SI [Chen-Wang-Jafar[14] & Wei-Banawan-Ulukus[15]]

- The user possesses S full messages in the cache.
- The user applies the classical PIR scheme.
- The user encodes the original queries by an MDS code.
- Maps the original queries into fewer queries by exploiting S messages.
- In both cases (fully and partially unknown), the capacity is given by:

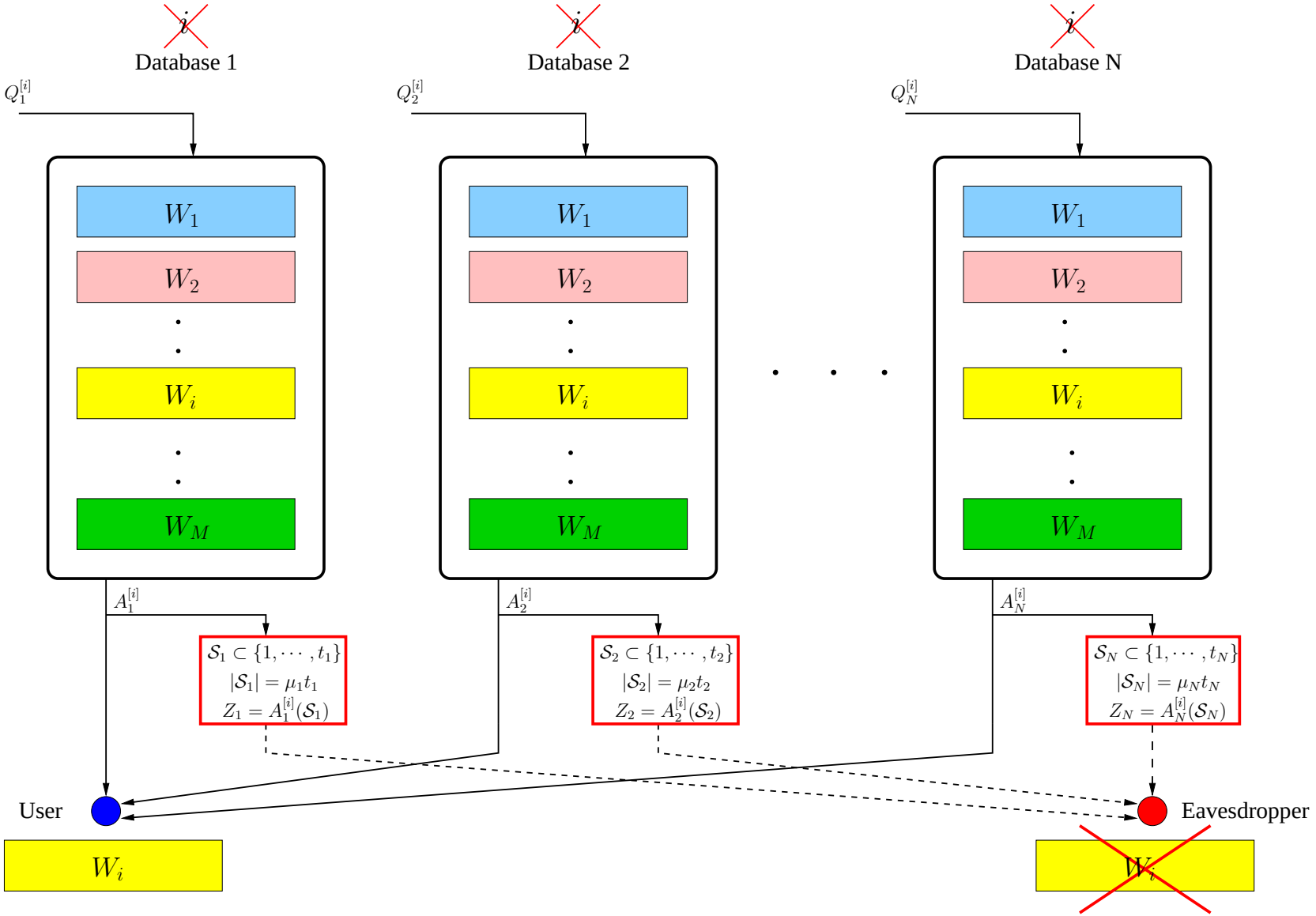
$$C = \frac{1 - \frac{1}{N}}{1 - (\frac{1}{N})^{M-S}}$$

- Partial knowledge of the side information does not hurt performance.

[14] Z. Chen, Z. Wang, and S. Jafar. The capacity of private information retrieval with private side information. 2017. Available at arXiv:1709.03022.

[15] Y.-P. Wei, K. Banawan, and S. Ulukus. The capacity of private information retrieval with partially known private side information. 2017. Available at arXiv:1710.00809.

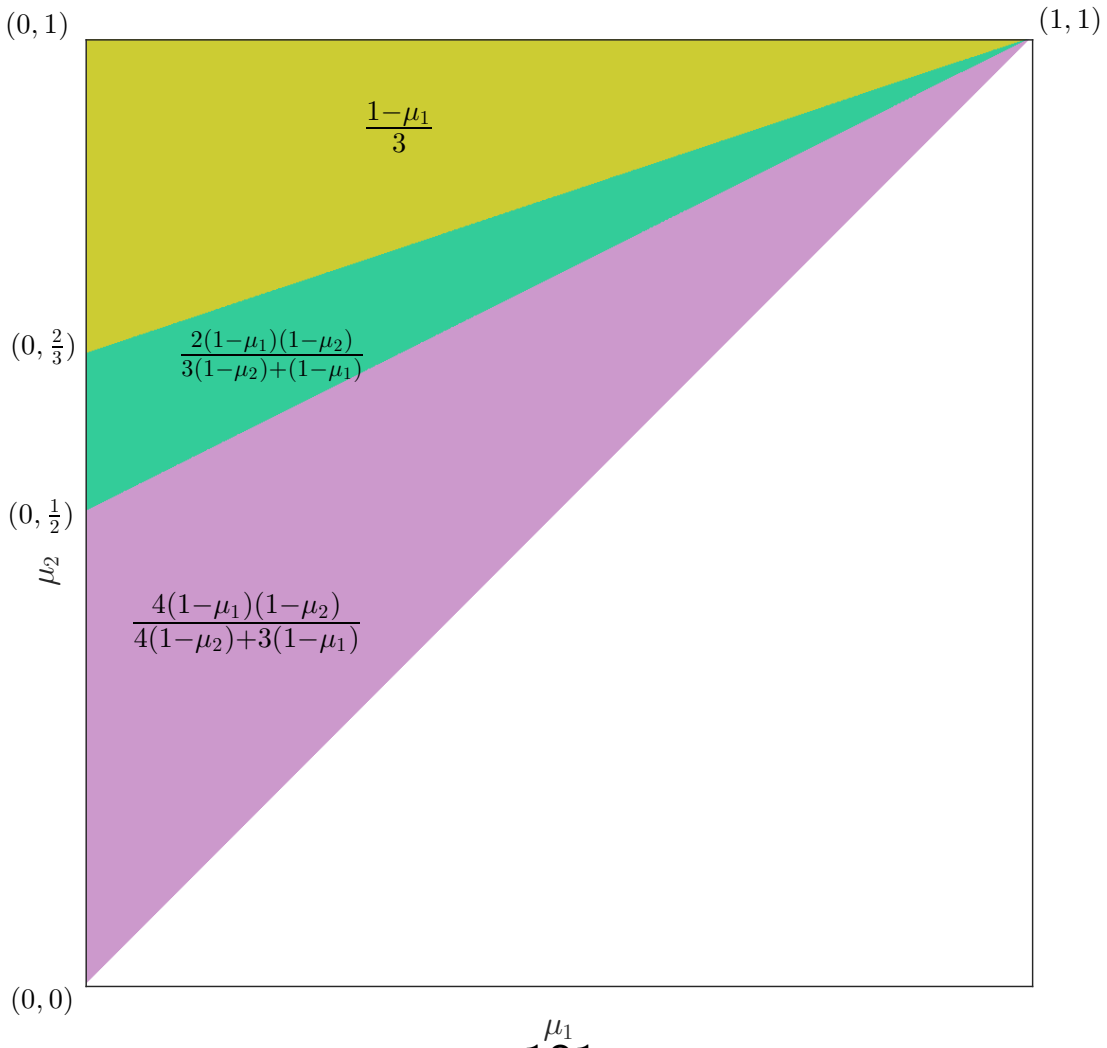
PIR over Wiretap Channel II [Banawan-Ulukus][16]



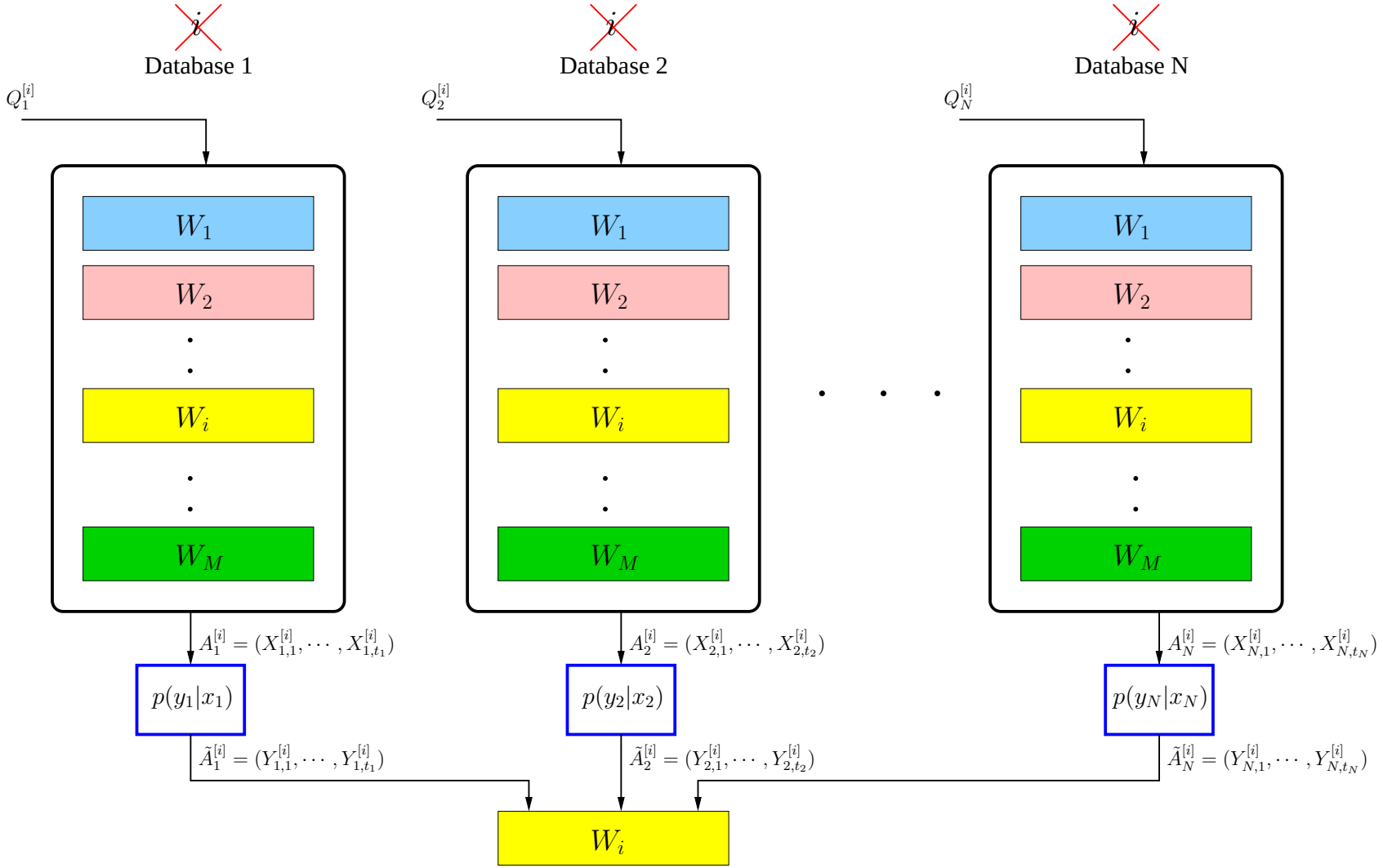
[16] K. Banawan, and S. Ulukus. Private Information Retrieval Through Wiretap Channel II: Privacy Meets Security. 2018. Available at arXiv:1801.06171.

Capacity Result for $M = 3$, $N = 2$

- The answers are encrypted to satisfy the security constraint.
- The answers are asymmetric in length.
- Different capacity expressions depending on (μ_1, μ_2).



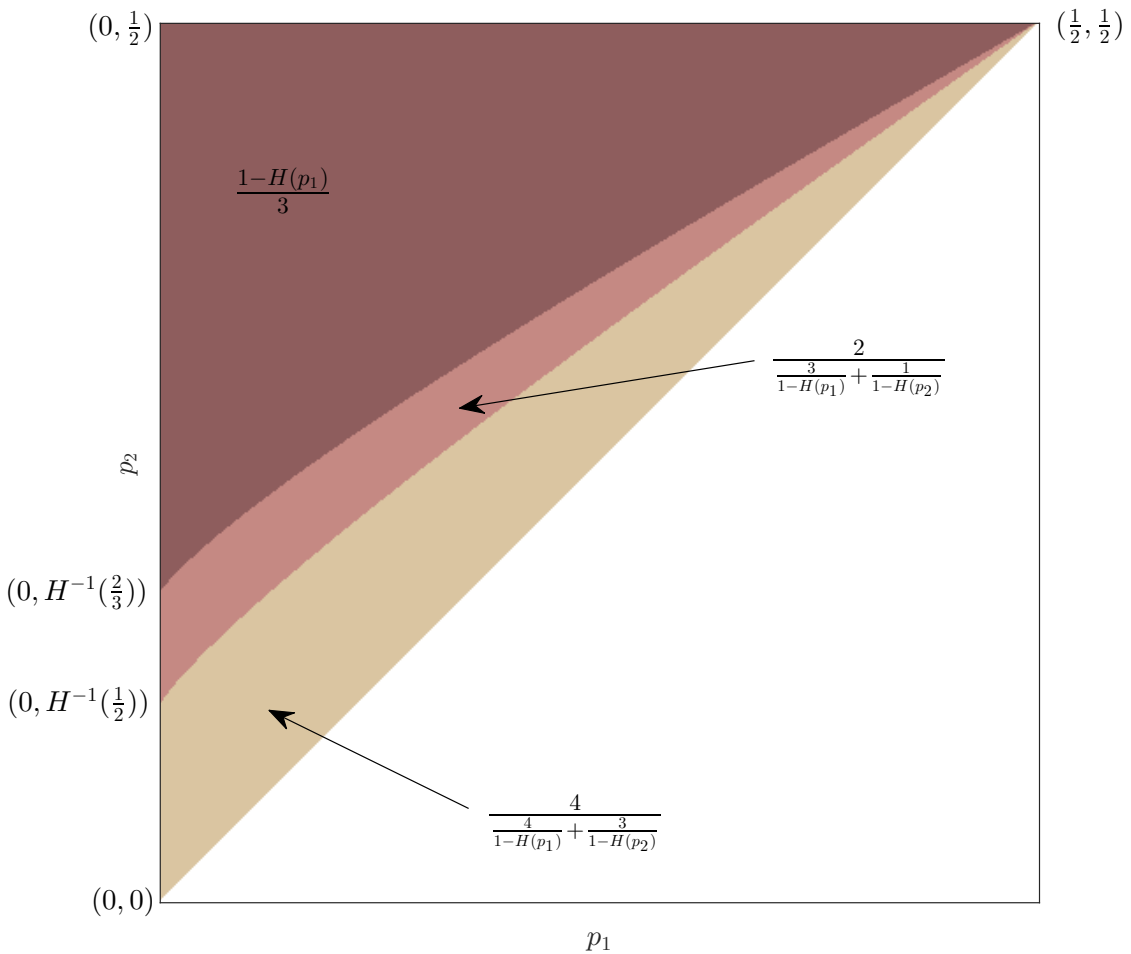
Noisy PIR (NPIR) [Banawan-Ulukus][17]



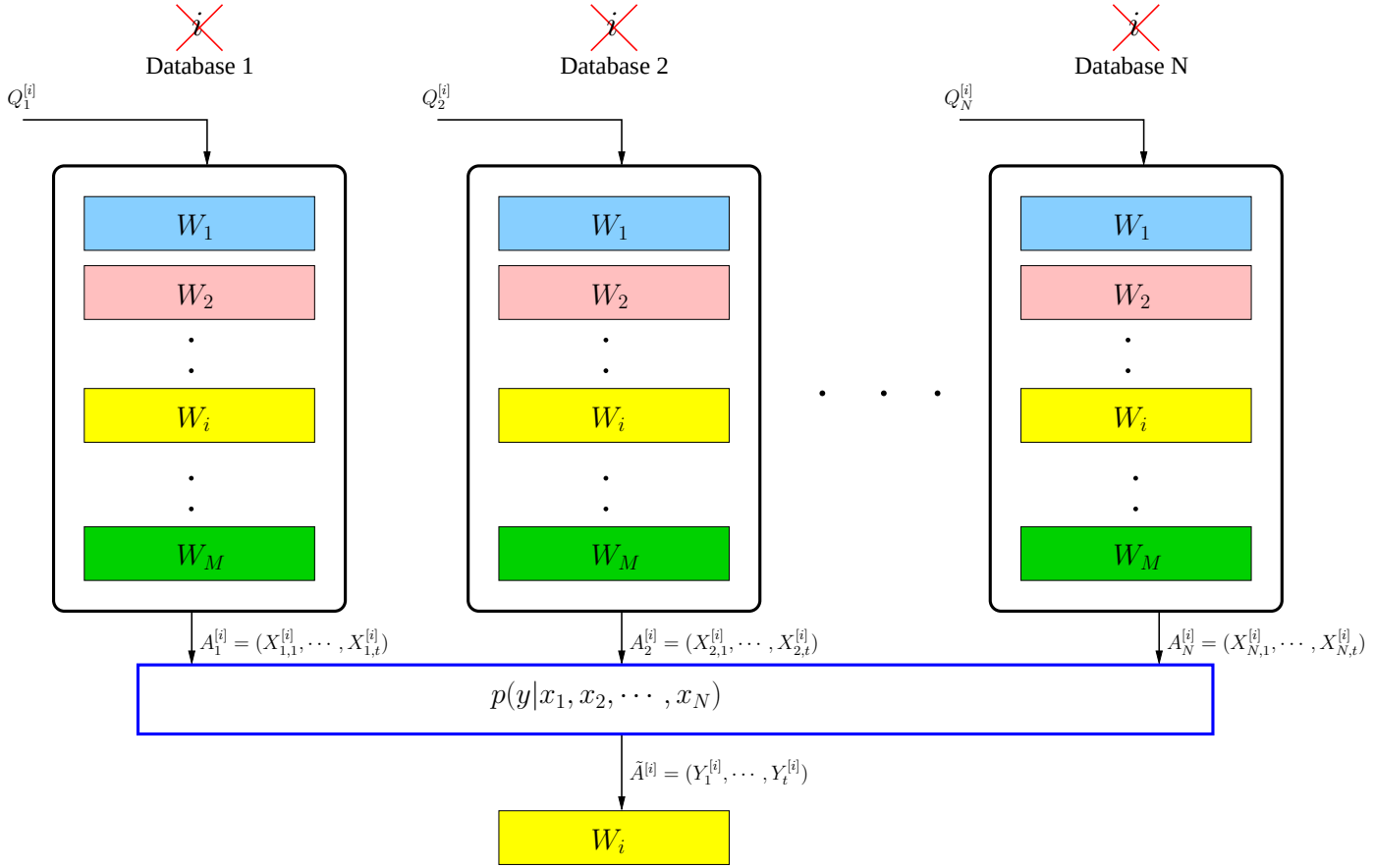
[17] K. Banawan, and S. Ulukus. Noisy Private Information Retrieval: On Separability of Channel Coding and Information Retrieval. 2018. Available at arXiv:1807.05997.

Capacity Result for $M = 3$, $N = 2$ from BSC(p_1) and BSC(p_2)

- Channel coding and the retrieval scheme are almost separable.
- Noisy channels affect only the traffic ratio requested from each database.
- Different capacity expressions depending on (p_1, p_2).
- Depends only on the capacity of channels not on transition probabilities.



PIR from Multiple Access Channels (MAC-PIR) [Banawan-Ulukus][17]



- Channel coding and the retrieval scheme are inseparable.
- Full unconstrained capacity may be attainable for some MACs.

[17] K. Banawan, and S. Ulukus. Noisy Private Information Retrieval: On Separability of Channel Coding and Information Retrieval. 2018. Available at arXiv:1807.05997.

Capacity Results: Summary

Model	Work of	PIR capacity
Classical PIR	Sun-Jafar	$\frac{1-\frac{1}{N}}{1-(\frac{1}{N})^M}$
Colluding (TPIR)	Sun-Jafar	$\frac{1-\frac{T}{N}}{1-(\frac{T}{N})^M}$
Robust-colluding (RPIR)	Sun-Jafar	$\frac{1-\frac{T}{N-U}}{1-(\frac{T}{N-U})^M}$
Symmetric (SPIR)	Sun-Jafar	$\begin{cases} 1-\frac{1}{N}, & \rho \geq \frac{1}{N-1} \\ 0, & \text{otherwise} \end{cases}$
Coded (CPIR)	Banawan-Ulukus	$\frac{1-\frac{K}{N}}{1-(\frac{K}{N})^M}$
Multi-message (MPIR)	Banawan-Ulukus	$\begin{cases} \frac{1}{1+\frac{M-P}{PN}}, & P \geq \frac{M}{2} \\ \frac{1-\frac{1}{N}}{1-(\frac{1}{N})^{M/P}}, & \frac{M}{P} \in \mathbb{N} \end{cases}$
Byzantine-colluding (BPIR)	Banawan-Ulukus	$\frac{N-2B}{N} \cdot \frac{1-\frac{T}{N-2B}}{1-(\frac{T}{N-2B})^M}$
Cache-aided (known-arbitrary)	Tandon	$\frac{1}{1-r} \cdot \frac{1-\frac{1}{N}}{1-(\frac{1}{N})^M}$
Cache-aided (unknown/partial-uncoded)	Wei-Banawan-Ulukus	low/high caching ratios
PIR with PSI (unknown/partial-full)	Chen-Wang-Jafar Wei-Banawan-Ulukus	$\frac{1-\frac{1}{N}}{1-(\frac{1}{N})^{M-S}}$
Asymmetric traffic	Banawan-Ulukus	upper and lower bounds
Wiretapped PIR	Banawan-Ulukus	upper and lower bounds

Conclusion

- **PIR schemes for:**
  - Classical PIR.
  - Colluding PIR.
  - Robust PIR.
  - Symmetric PIR.

- **Our contributions:**
  - Coded PIR.
  - Multi-message PIR.
  - Byzantine PIR.
  - PIR under asymmetric traffic constraints.
  - Cache-aided PIR.
  - PIR from WTC-II.
  - Noisy PIR.
  - PIR from MACs.

- Many open problems $\cdots$