Enhancing Extended Reality Experience with Location-Dependent Multiantenna Coded Caching

Antti Tölli
with Hamidreza Bakhshzad Mahmoodi and MohammadjavadSalehi
e-mail: antti.tolli@oulu.fi

Centre for Wireless Communications
University of Oulu

- M. Salehi, K. Hooli, J. Hulkkonen and A. Tölli, "Enhancing Next-Generation Extended Reality Applications with Coded Caching," in IEEE Open Journal of the Communications Society, doi: 10.1109/OJCOMS.2023.3286473, 2023.
- H. B. Mahmoodi, M. Salehi and A. Tölli, "Multi-antenna Coded Caching for Location-Dependent Content Delivery," in IEEE Transactions on Wireless Communications, doi: 10.1109/TWC.2023.3277983, 2023.
- H. B. Mahmoodi, M. Salehi and A. Tölli, "Low-Complexity Multi-Antenna Coded Caching Using Location-Aware Placement Delivery Arrays," submitted to IEEE TWC, arXiv preprint arXiv:2305.06858 (2023).

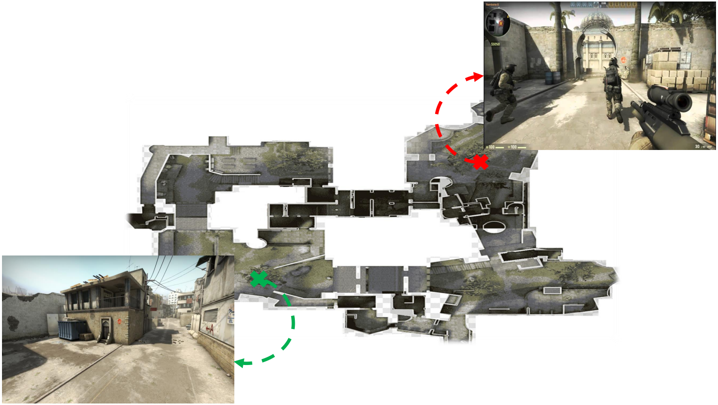
UNIVERSITY OF OULU

Multi-user XR

Users immersed into the interactive virtual world via 3D interface

Highly reliable high-throughput wireless connections with low latency are required.

Location-Dependent Multimedia Content

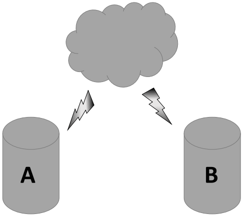


Large part of static (infrastructure) and dynamic (avatars, shapes) content can be stored in advance at user devices.

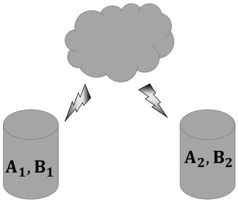
[1] M. Salehi, K. Hooli, J. Hulkkonen and A. Tölli, "Enhancing Next-Generation Extended Reality Applications with Coded Caching," in IEEE Open Journal of the Communications Society, doi: 10.1109/OJCOMS.2023.3286473.

Coded Caching (CC)

Traditional Caching



Coded Caching

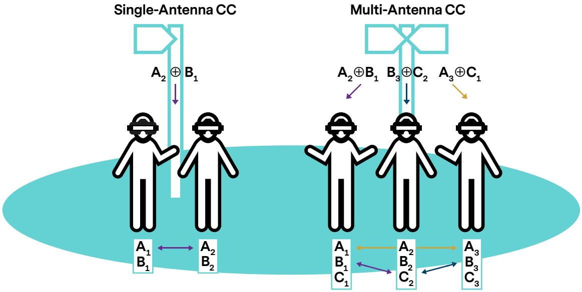


Request 1	Request 2	Load
A	B	0
B	A	2
A	A	1
B	B	1
Average load		1

Request 1	Request 2	Codeword	Load
A	B	$A_2 \oplus B_1$	0.5
B	A	$B_2 \oplus A_1$	0.5
A	A	$A_2 \oplus A_1$	0.5
B	B	$B_2 \oplus B_1$	0.5
Average load			0.5

CC gain is proportional to the total cache size in the network.

Multi-Antenna Coded Caching



CC gain is additive with the spatial multiplexing gain.

[2] S. P. Shariatpanahi, G. Caire, and B. Hossein Khalaj, "Physical-layer schemes for wireless coded caching," IEEE Transactions on Information Theory, vol. 65, no. 5, pp. 2792–2807, May 2019.

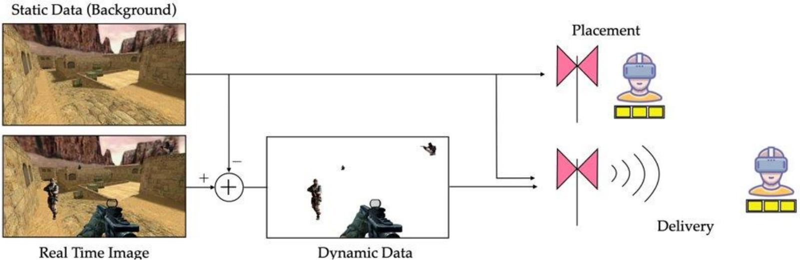
[3] A. Tölli, S. P. Shariatpanahi, J. Kaleva and B. H. Khalaj, "Multi-Antenna Interference Management for Coded Caching," in IEEE Transactions on Wireless Communications, vol. 19, no. 3, pp. 2091-2106, March 2020

[4] M. Salehi, E. Parrinello, S. P. Shariatpanahi, P. Elia and A. Tölli, "Low-Complexity High-Performance Cyclic Caching for Large MISO Systems," in IEEE Transactions on Wireless Communications, vol. 21, no. 5, pp. 3263-3278, May 2022,

Location-dependent (Coded) Caching [5, 6]

Location dependency of the content to alleviate wireless bottlenecks:

- Use a priori knowledge of the expected throughput per location;
- Allocate higher cache share for content relevant to locations with poor connectivity;

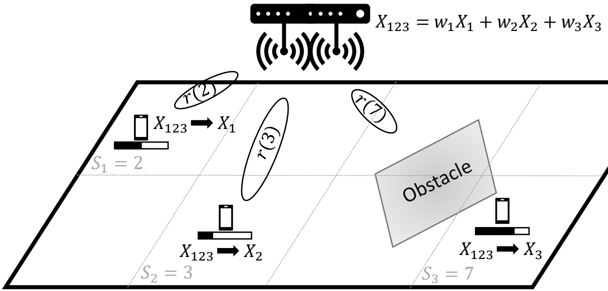


Can we enable a CC gain for such a non-uniform memory allocation?

[5] H. B. Mahmoodi, M. Salehi and A. Tölli, "Multi-antenna Coded Caching for Location-Dependent Content Delivery," in IEEE Transactions on Wireless Communications, doi: 10.1109/TWC.2023.3277983 (2023).

[6] H. B. Mahmoodi, M. Salehi and A. Tölli, "Low-Complexity Multi-Antenna Coded Caching Using Location-Aware Placement Delivery Arrays," submitted to IEEE TWC, arXiv preprint arXiv:2305.06858 (2023).

System Model



- Dynamic environment with K users.
- Coverage area divided into $|\mathcal{S}|$ single transmission units (STU).
- Memory allocation per STU based on predicted/expected rate, e.g.:

$$r(s) = \frac{\Omega}{F} C_p \mathbb{E}\left[\log(1 + \frac{P_T \, \|\mathbf{h}_{k_s}\|^2}{N_0})\right] \quad \text{[files/second]} \qquad (1)$$

UNIVERSITY OF OULU

Memory Allocation

- Placement phase: 1) memory allocation and 2) cache placement.
- Worst case total delivery time is first approximated based on $r(s), \forall s \in \mathcal{S}$:

$$\hat{T}_T = \frac{K}{K\bar{m} + L} \max_{s \in \mathcal{S}} \frac{1 - m(s)}{r(s)} \quad \text{[seconds]}, \tag{2}$$

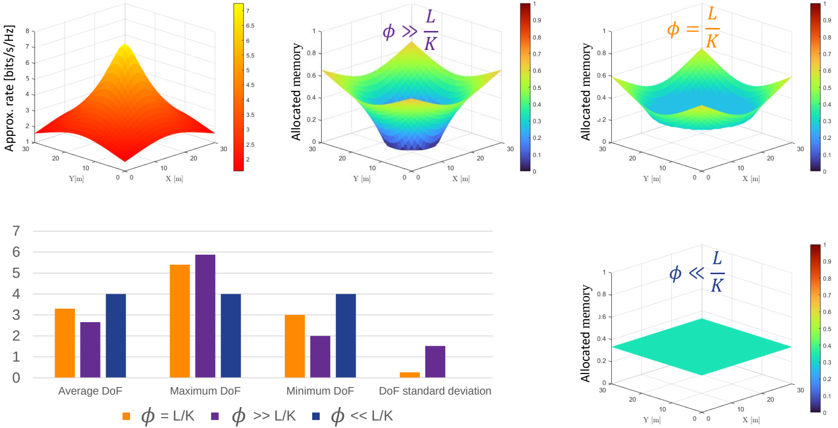
where $\bar{m} = \min_{s \in \mathcal{S}} m(s)$.

- Memory allocation $m(s)\ \forall\ s$ is performed based on minimizing (2):

$$\begin{aligned} &\min_{m(s),\, \gamma \geq 0,\, \bar{m} \geq 0} \quad \frac{\gamma}{\bar{m} + \frac{L}{K}} \\ &\text{s.t.} \quad \frac{1 - m(s)}{r(s)} \leq \gamma,\ \forall s \in \mathcal{S}, \\ &\bar{m} \leq m(s),\ \forall s \in \mathcal{S}, \quad \sum_{s \in \mathcal{S}} m(s) \leq M. \end{aligned} \tag{3}$$

Memory Allocation – Example

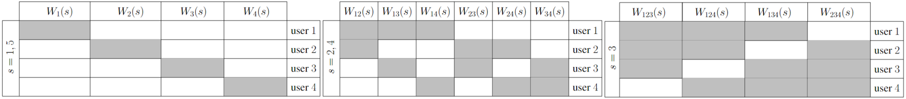
- The term $\frac{L}{K}$ in (3) can be substituted with a general parameter ϕ.
- A $30 \times 30[m^2]$ room with $K = 6$ users, $L = 2$ transmit antennas, and $\frac{M}{S} = 0.33$ relative memory size.



Cache placement

- Cache placement can be based, e.g., on multi-server scheme [5] or PDA-based scheme [6].
- Non-uniform placement of STU-specific file-fragments.
- A toy example with 5 STUs [5]:

	$s=1$	$s=2$	$s=3$	$s=4$	$s=5$
$r(s)$	3×10^3	2×10^3	1×10^3	2×10^3	3×10^3
$m(s)$	0.25	0.5	0.75	0.5	0.25

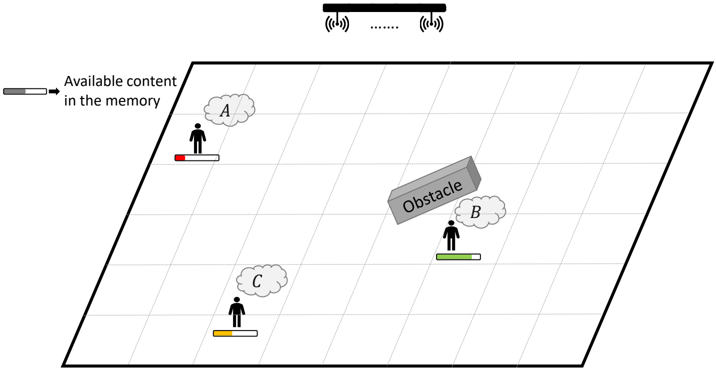


[5] H. B. Mahmoodi, M. Salehi and A. Tölli, "Multi-antenna Coded Caching for Location-Dependent Content Delivery," in IEEE Transactions on Wireless Communications, doi: 10.1109/TWC.2023.3277983 (2023).

[6] H. B. Mahmoodi, M. Salehi and A. Tölli, "Low-Complexity Multi-Antenna Coded Caching Using Location-Aware Placement Delivery Arrays," submitted to IEEE TWC, arXiv preprint arXiv:2305.06858 (2023).

Content Delivery

Delivery phase: users request location-dependent content.



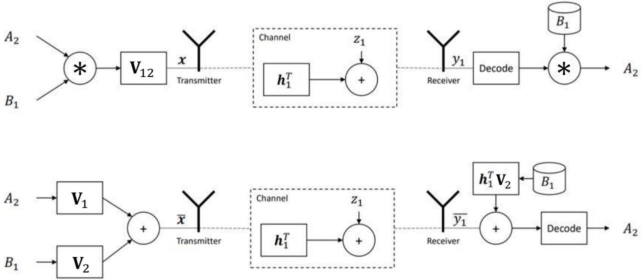
The BS transmits precoded messages $\mathbf{x}_{\mathcal{U}}$ to users $\mathcal{U} \subseteq \mathcal{K}$ over several transmit intervals.

UNIVERSITY OF OULU

Different Delivery Methods

- NCM-based scheme [5] enables parallel multicast transmissions to user groups with size $|\mathcal{T}|$: $\mathbf{x}_{\mathcal{U}} = \sum_{\mathcal{T} \subseteq \mathcal{U}} \mathbf{v}_{\mathcal{T}} x_{\mathcal{T}}$.
- The PDA-based scheme [6] enables unicast transmission by sending $\mathbf{x}_{\mathcal{U}} = \sum_{k \in \mathcal{U}} \mathbf{v}_k x_k$.
- Achievable degrees of freedom (DoF) dor both schemes is $\min_{k \in \mathcal{U}} Km(s_k) + L$.

$$(A_2 * B_1)\mathbf{V}_{12} \quad \rightarrow \quad A_2\mathbf{V}_1 + B_1\mathbf{V}_2$$



UNIVERSITY OF OULU

Different Transmission/Reception Signals (PDA - based)

- Consider the following transmitted signal using PDA,

$$\mathbf{x}_{\mathsf{PDA}}(1) = A_4\mathbf{v}_1 + B_1\mathbf{v}_2 + C_1\mathbf{v}_3$$

- Corresponding received signal at the users,

$$\begin{aligned} y_1 &= A_4\mathbf{h}_1^H\mathbf{v}_1 + \underline{B_1}\mathbf{h}_1^H\mathbf{v}_2 + \underline{C_1}\mathbf{h}_1^H\mathbf{v}_3 + w_1, \\ y_2 &= \underline{A_4}\mathbf{h}_2^H\mathbf{v}_1 + B_1\mathbf{h}_2^H\mathbf{v}_2 + \underline{\underline{C_1}}\mathbf{h}_2^H\mathbf{v}_3 + w_2, \\ y_3 &= \underline{\underline{A_4}}\mathbf{h}_3^H\mathbf{v}_1 + \underline{\underline{B_1}}\mathbf{h}_3^H\mathbf{v}_2 + C_1\mathbf{h}_3^H\mathbf{v}_3 + w_3 \end{aligned}$$

- Underlined terms are regenerated and removed using cache and the double-underlined terms suppressed by precoder $\mathbf{v}_k$.
- No need for SIC at the receiver.

Weighed Maxmin (WMM) Beamforming

- Different fractions of files missing given the UE locations $s_k, k \in \mathcal{U}$.
- The WMM multicast beamformer design [5] (l.h.s) and the unicast beamformer design [6] (r.h.s):

$$\max_{\{\mathbf{v}_\mathcal{T}, \gamma_\mathcal{T}^k, R_{\text{sum}}^\mathcal{Q}\}} \min_{k \in \mathcal{U}} \min_{\mathcal{Q} \subseteq \mathcal{D}_k} \frac{1}{c_k|\mathcal{Q}|} R_{\text{sum}}^\mathcal{Q}$$

$$\text{s. t.} \quad R_{\text{sum}}^\mathcal{Q} \leq \log\left(1 + \sum_{\mathcal{T} \in \mathcal{Q}} \gamma_\mathcal{T}^k\right),$$

$$\gamma_\mathcal{T}^k \leq \frac{|\mathbf{h}_k^H \mathbf{v}_\mathcal{T}|^2}{\sum_{\mathcal{V} \in \mathcal{I}_k} |\mathbf{h}_k^H \mathbf{v}_\mathcal{V}|^2 + \sigma^2},$$

$$\sum_{\mathcal{T} \subseteq \mathcal{U}, |\mathcal{T}| = \hat{t}+1} \|\mathbf{v}_\mathcal{T}\|^2 \leq P_T,$$

$$\max_{\{\mathbf{v}_k, \gamma_k\}} \min_{k \in \mathcal{U}} \frac{1}{c_k} \log(1 + \gamma_k)$$

$$\text{s.t.}$$

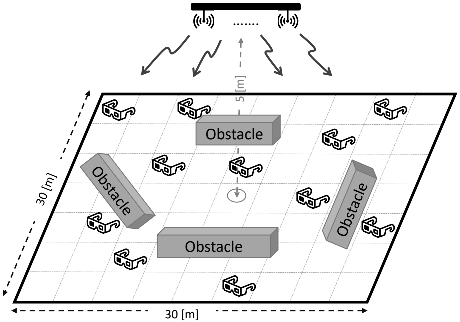
$$\gamma_k \leq \frac{|\mathbf{h}_k^H \mathbf{v}_k|^2}{\sum_{i \in \overline{\mathcal{I}}_k} |\mathbf{h}_k^H \mathbf{v}_i|^2 + N_0},$$

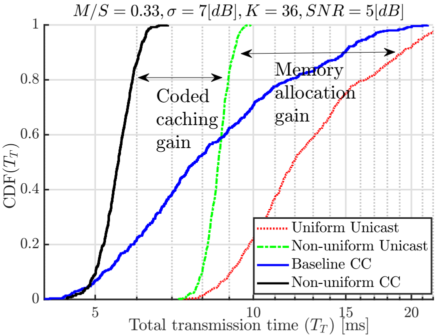
$$\sum_{k \in \mathcal{U}} \|\mathbf{v}_k\|^2 \leq P_T,$$

Comparison with uniform placement: lower DoF $\frac{(K \min_k(m_k) + L)}{(K\frac{M}{S} + L)} < 1$ is compensated by improved transmission rates $\frac{R_w}{R_u} > 1$.

Numerical Examples [6]

Large number of users ($K = 12 - 60$). Parameter σ models the impact of random placement of obstacles.

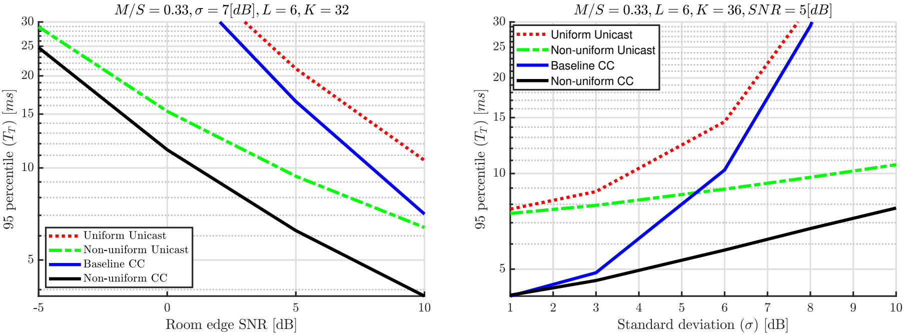




Significant performance improvement due to optimal memory allocation $\rightarrow$ further gain by leveraging coded caching transmission.

UNIVERSITY OF OULU

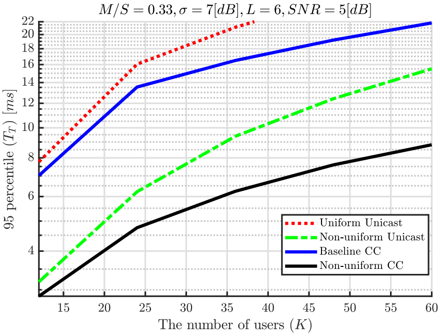
Simulation Results [6]

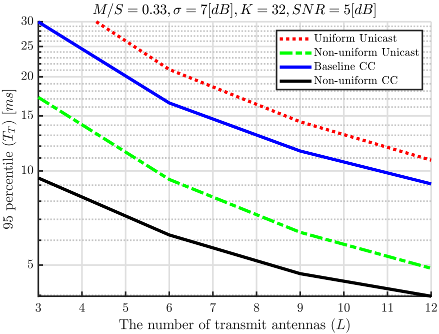


- Higher gains at lower available transmit powers.
- At very high SNR, non-uniform placement provides less gains.
- Largest gains available when pathloss variation (σ) is high.

[6] H. B. Mahmoodi, M. Salehi and A. Tölli, "Low-Complexity Multi-Antenna Coded Caching Using Location-Aware Placement Delivery Arrays,"submitted to IEEE TWC, arXiv preprint arXiv:2305.06858 (2023).

Simulation Results [6]





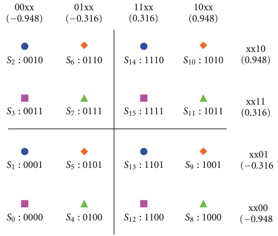
- Higher achievable coded caching gain $K \min m(s_k)$ in larger networks $\rightarrow$ higher performance gain.
- When the number of spatial multiplexing gain L is relatively larger than CC gain, unicast transmission achieves almost the same performance as CC-based schemes.

UNIVERSITY OF OULU

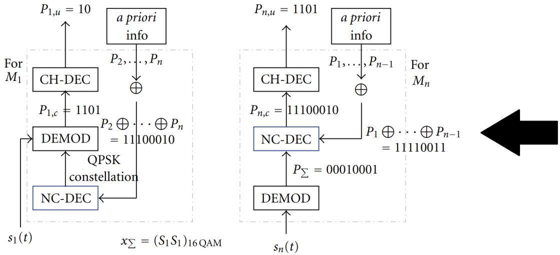
Conclusion

- Location-aware CC scheme is proposed for location-dependent data delivery to avoid excessive transmission time.
- A trade-off between local and global caching gain is achieved through novel memory allocation.
- Using the CC scheme a global caching gain additive to the multiplexing gain is achieved.
- Most effective in scenarios with large expected throughput variation.

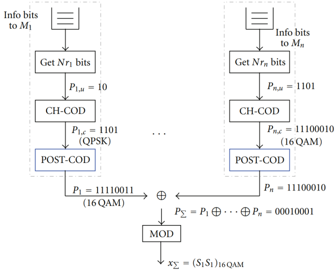
Nested Code Modulation



- Multi-rate transmission, based on side information at the RX
- An example of nesting QPSK constellations within 16 QAM .



Recover information bits



Coding and modulation

Different Transmission/Reception Signals (NCM - based)

- Consider the following transmitted signal using NCM,

$$\mathbf{x}_{\mathsf{NCM}}(1) = [A_2 * B_1]\mathbf{v}_{12} + [A_3 * C_1]\mathbf{v}_{13} + [B_3 * C_2]\mathbf{v}_{23}$$

- Corresponding received signal at the users,

$$y_1 = [A_2 * \underline{B_1}]\mathbf{h}_1^H\mathbf{v}_{12} + [A_3 * \underline{C_1}]\mathbf{h}_1^H\mathbf{v}_{13} + \underline{\underline{[B_3 * C_2]}}\mathbf{h}_1^H\mathbf{v}_{23} + w_1,$$
$$y_2 = [\underline{A_2} * B_1]\mathbf{h}_2^H\mathbf{v}_{12} + \underline{\underline{[A_3 * C_1]}}\mathbf{h}_2^H\mathbf{v}_{13} + [B_3 * \underline{C_2}]\mathbf{h}_2^H\mathbf{v}_{23} + w_2,$$
$$y_3 = \underline{\underline{[A_2 * B_1]}}\mathbf{h}_3^H\mathbf{v}_{12} + [\underline{A_3} * C_1]\mathbf{h}_3^H\mathbf{v}_{13} + [\underline{B_3} * C_2]\mathbf{h}_3^H\mathbf{v}_{23} + w_3$$

- Underlined terms are removed using cache and the double-underlined terms are suppressed using precoder $\mathbf{v}_{ij}$.
- Superposition of multiple messages at each user → successive interference cancellation (SIC) at UEs.