The Influence of CSI in Ultra-Reliable Low-Latency Communications with IR-HARQ

Apostolos Avranas, Marios Kountouris, Philippe Ciblat

INSTITUT POLYTECHNIQUE DE PARIS

HUAWEI

Supporting New Applications:

Mission-critical control

smart grid (5ms for tx/grid backbone, 50ms for grid backhaul)

Industrial Automation

real time control (20µs – 10 ms), time-critical sensing (10ms), over-the-air: 0.25~1 ms

Media & Entertainment

4K Video, live streaming in crowded areas (20ms), collaborative gaming (20ms)

AR/VR

MTP 15~20 ms // over-the-air:1~5 ms

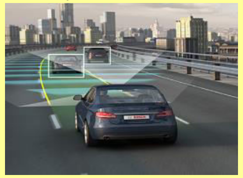
Tactile Internet

1ms e2e latency

HF Trading & eCommerce

1ms delay advantage = 100M$/year

A page load slowdown of 1 sec could cost 1.6B$ in sales each year (Amazon)



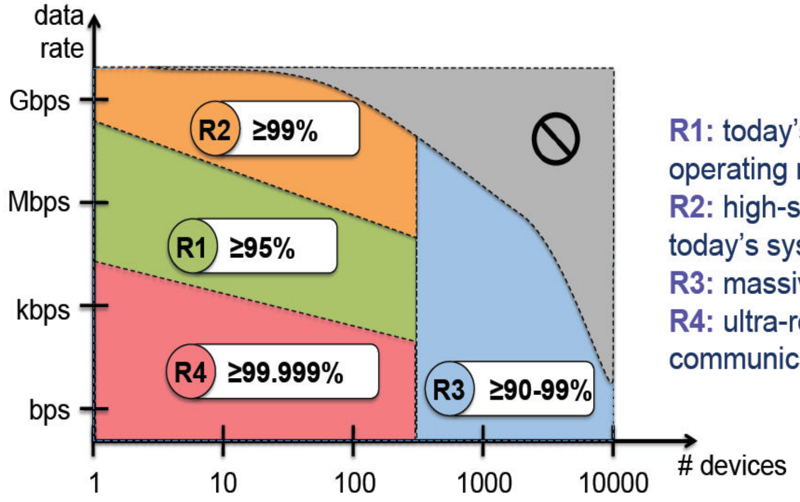
Automotive industry

autonomous/cooperative driving, V2X (3-10ms), V2N for remote vehicle operation (10-30ms E2E)

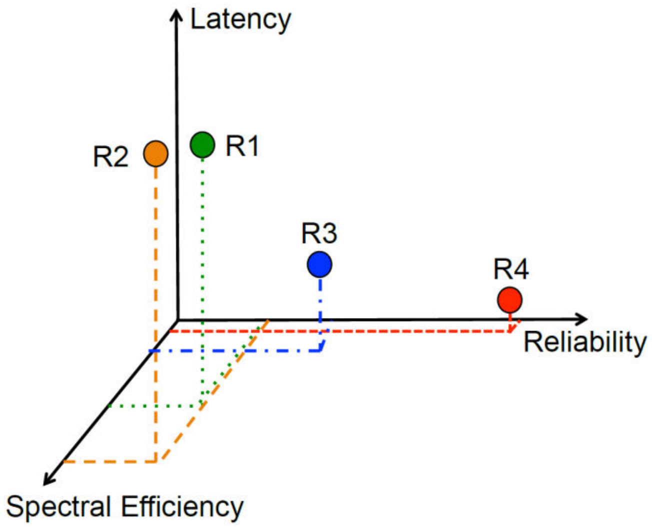
Healthcare

video/AR for remote surgery (100ms), real-time command/control (10-100ms)

New Operation Regimes – URLLC



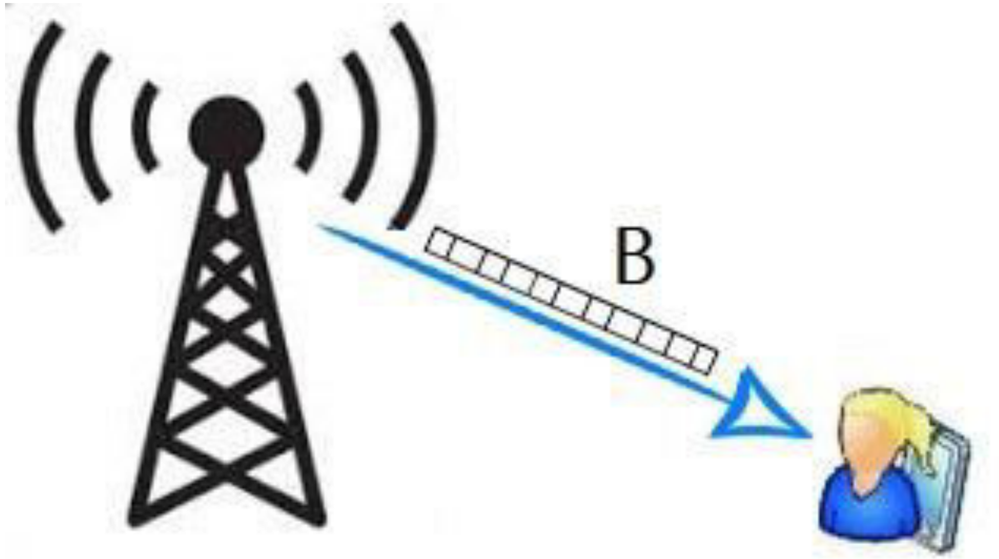
R1: today's systems operating region
R2: high-speed versions of today's systems
R3: massive access
R4: ultra-reliable communication



From F. Boccardi et al. "Five Disruptive Technology Directions for 5G

Use case	Latency (ms)	Reliability (PEP)	Data size (bytes)	Communication range (m)
Industrial automation	0.25 – 10	$10^{-6} - 10^{-9}$	10 – 300	10 – 100
Smart grids	5 – 50	10^{-6}	80 – 1000	Few m to km
Intelligent transport systems	5 – 100	$10^{-3} - 10^{-5}$	500 – 1k	200 – 5000
Telemedicine	1 – 10 (haptics) 20 – 100 (video, audio)	10^{-5}	200 – 4k	$<$ 200 km

System model



- Point-2-Point communication
- Fixed number ***B*** of information bits
- Messaged transmitted using maximum ***N*** channel uses
- Transmissions within time coherence interval, i.e. constant channel gain ***g***

Error Probability: $\mathcal{E}$

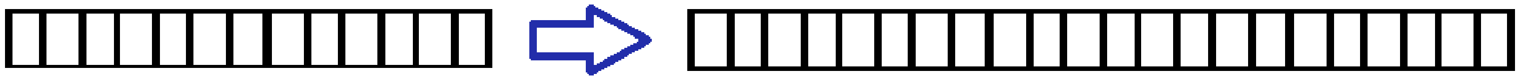
- If block-length $N \to \infty$ and mutual information above a threshold then:
  $\varepsilon \to 0$

Error Probability: $\mathcal{E}$

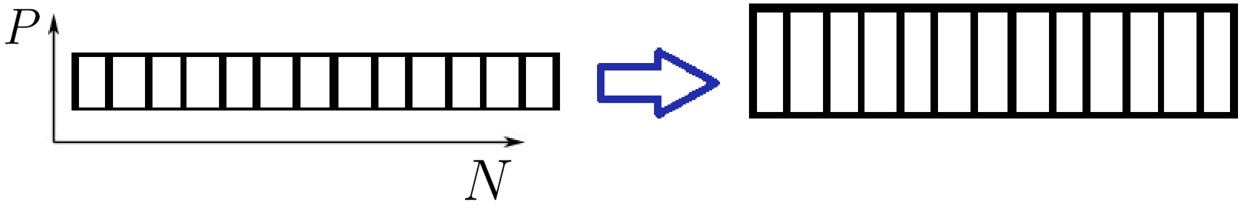
- If block-length $N \to \infty$ and mutual information is above a threshold: $\varepsilon \to 0$
- If block-length finite: $\varepsilon > 0$. Two straightforward ways of improving:

Error Probability: $\mathcal{E}$

- If block-length $N \to \infty$ and mutual information above a threshold then: $\mathcal{E} \to 0$
- If block-length finite: $\mathcal{E} > 0$. Two straightforward ways of improving:
  - Increase N :



  - Increase P :



Error Probability: $\mathcal{E}$

$$\varepsilon \approx Q\left(\frac{N\ln(1+gP) - B\ln 2 + \frac{\ln N}{2}}{\sqrt{N(1-\frac{1}{(1+gP)^2})}}\right)$$

[1] Y. Polyanskiy, "Channel coding: Non-asymptotic fundamental limits", Ph.D. dissertation, Princeton University, Nov. 2010.

[2] M. Hayashi, "Information spectrum approach to second-order coding rate in channel coding," IEEE Trans. on Inf. Theory, vol. 55, no. 11, pp. 4947–4966, Nov. 2009

Error Probability: $\mathcal{E}$

$$\varepsilon \approx Q\left(\frac{N\ln(1+gP) - B\ln 2}{\sqrt{N(1-\frac{1}{(1+gP)^2})}}\right)$$

[1] Y. Polyanskiy, “Channel coding: Non-asymptotic fundamental limits”, Ph.D. dissertation, Princeton University, Nov. 2010.

[2] M. Hayashi, “Information spectrum approach to second-order coding rate in channel coding,” IEEE Trans. on Inf. Theory, vol. 55, no. 11, pp. 4947–4966, Nov. 2009

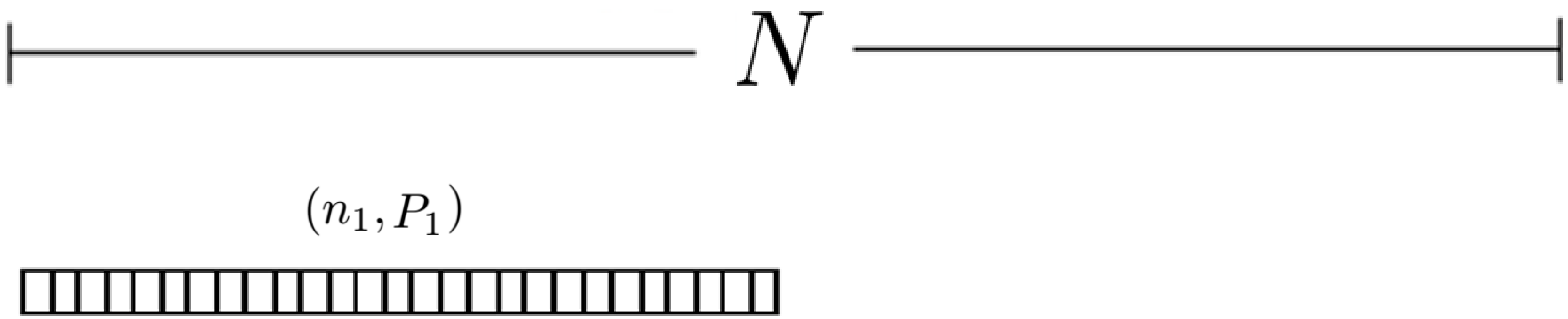
Still: $\varepsilon > 0$

- Want to keep it small without always spending too much (N, P)

Still: $\varepsilon > 0$

- Want to keep it small without always spending too much (N, P)

- Incremental Redundancy Hybrid Automatic Repeat Request (IR-HARQ). For $M = 2$ rounds:



Still: $\varepsilon > 0$

- Want to keep it small without always spending too much (N, P)

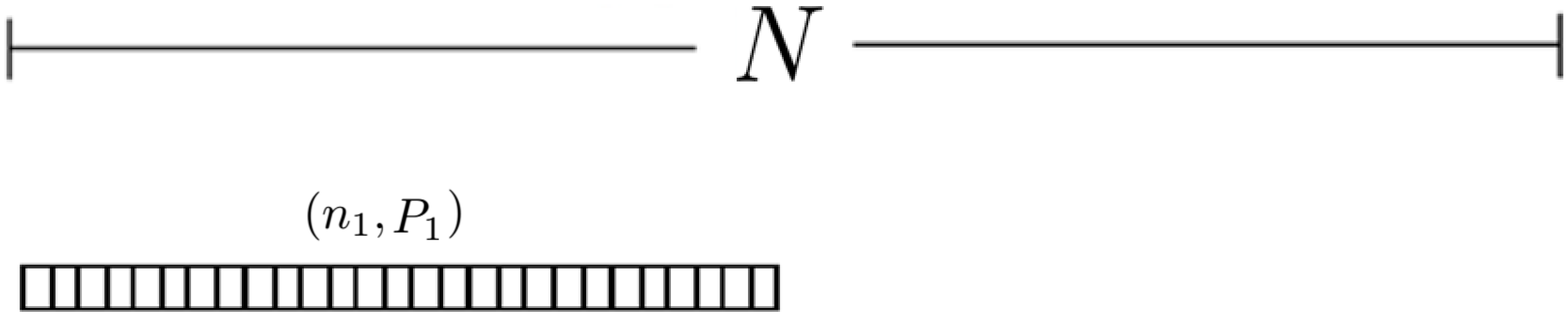
- Incremental Redundancy Hybrid Automatic Repeat Request (IR-HARQ). For $M = 2$ rounds:

(n_1, P_1)

Still: $\varepsilon > 0$

- Want to keep it small without always spending too much (N, P)

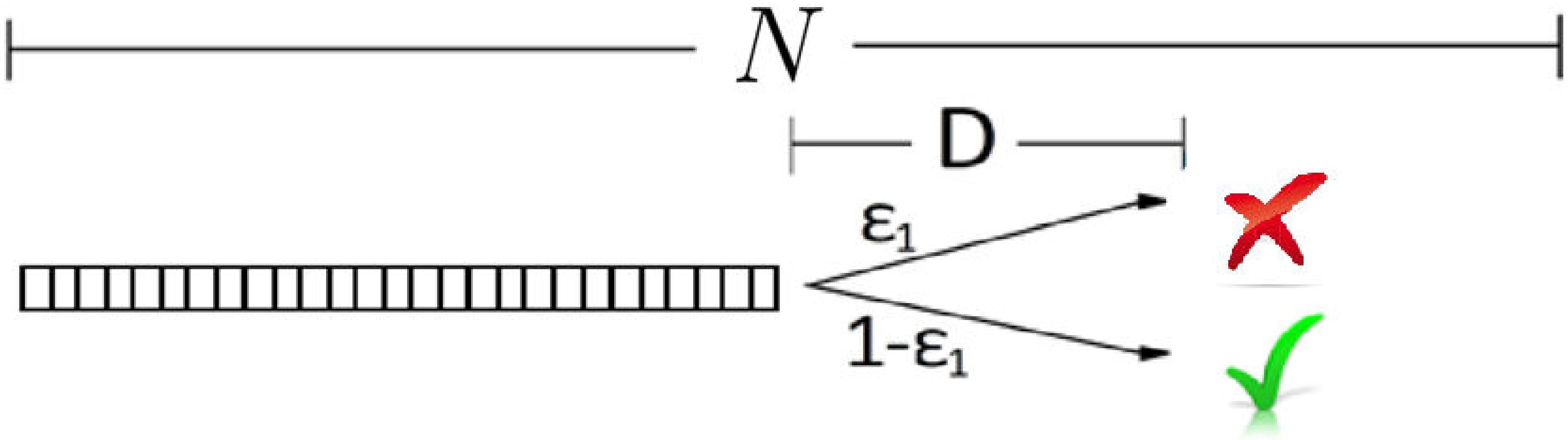
- Incremental Redundancy Hybrid Automatic Repeat Request (IR-HARQ). For $M = 2$ rounds:



Still: $\varepsilon > 0$

- Want to keep it small without always spending too much (N, P)

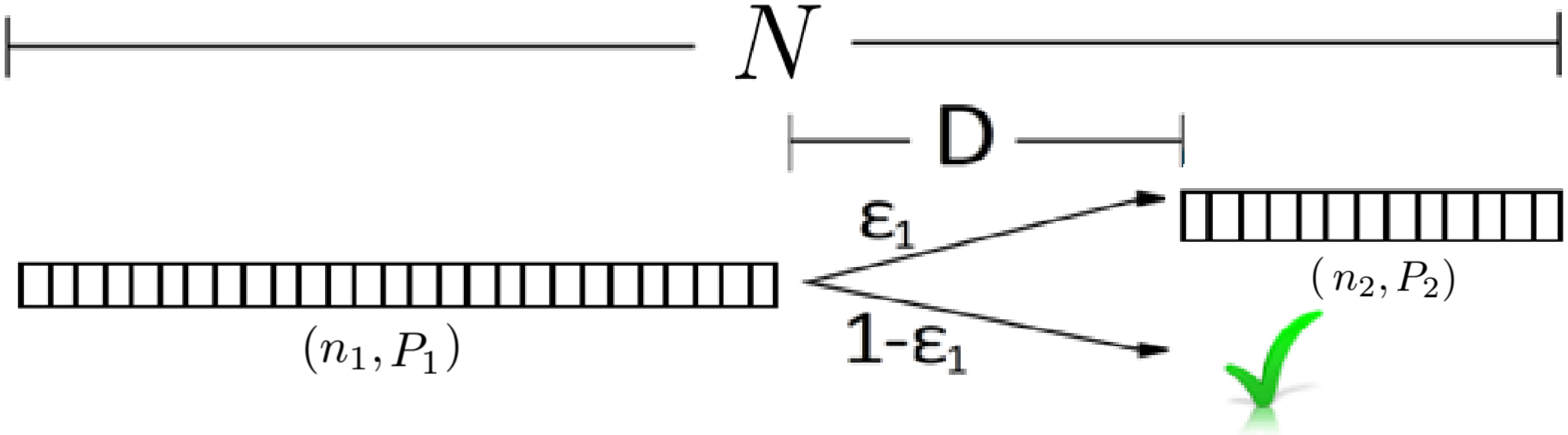
- Incremental Redundancy Hybrid Automatic Repeat Request (IR-HARQ). For $M = 2$ rounds:



Still: $\varepsilon > 0$

- Want to keep it small without always spending too much (N, P)

- Incremental Redundancy Hybrid Automatic Repeat Request (IR-HARQ). For $M = 2$ rounds:



Error Probability ε_m for IR-HARQ at m round

$$\varepsilon_m \approx Q\left(\frac{\sum_{i=1}^{m} n_i \ln(1+gP_i) - B\ln 2}{\sqrt{\sum_{i=1}^{m} n_i \left(1 - \frac{1}{(1+gP_i)^2}\right)}}\right)$$

Optimization Problem, URLLC,
Fixing: B information bits, M = 2, $D = 0$

$$\min_{n_1, n_2, P_1, P_2} \quad (\alpha\text{-}1)\text{Throughput} + \alpha\text{Energy}$$

$s.t.$ Reliability

Latency

Energy & Maximum Power

Optimization Problem, *Full-CSI*

Fixing: B information bits, M = 2, $D = 0$

$$\min_{n_1(g), n_2(g), P_1(g), P_2(g)} \mathbb{E}_g[-\frac{\mathcal{T}_h(a)}{\mathcal{T}_{h,max}} + \frac{\mathcal{E}(a)}{\mathcal{E}_{min}}]$$

Optimization Problem, *Full-CSI*

Fixing: B information bits, M = 2, $D = 0$

$$\min_{n_1(g),n_2(g),P_1(g),P_2(g)} \mathbb{E}_g[-\frac{\overbrace{\mathcal{T}_h(a)}^{(1-a)\left(\frac{B(1-\varepsilon_2)}{n_1+n_2\varepsilon_1}\right)}}{\mathcal{T}_{h,max}} + \frac{\overbrace{\mathcal{E}(a)}^{a\left(n_1P_1+n_2P_2\varepsilon_1\right)}}{\mathcal{E}_{min}}]$$

Optimization Problem, *Full-CSI*

Fixing: B information bits, M = 2, $D = 0$

$$\min_{n_1(g),n_2(g),P_1(g),P_2(g)} \mathbb{E}_g[-\frac{\overbrace{(1-a)\left(\frac{B(1-\varepsilon_2)}{n_1+n_2\varepsilon_1}\right)}^{\mathcal{T}_h(a)}}{\mathcal{T}_{h,max}} + \frac{\overbrace{a\left(n_1P_1+n_2P_2\varepsilon_1\right)}^{\mathcal{E}(a)}}{\mathcal{E}_{min}}]$$

$$s.t. \quad n_1(g)+n_2(g) \leq \mathrm{N}_\ell, \quad \forall g$$

Latency

Optimization Problem, *Full-CSI*

Fixing: B information bits, M = 2, $D = 0$

$$\min_{n_1(g), n_2(g), P_1(g), P_2(g)} \mathbb{E}_g\left[-\frac{\overbrace{\mathcal{T}_h(a)}^{(1-a)\left(\frac{B(1-\varepsilon_2)}{n_1+n_2\varepsilon_1}\right)}}{\mathcal{T}_{h,max}} + \frac{\overbrace{\mathcal{E}(a)}^{a\left(n_1P_1+n_2P_2\varepsilon_1\right)}}{\mathcal{E}_{min}}\right]$$

$$s.t. \quad n_1(g) + n_2(g) \leq \mathrm{N}_\ell, \quad \forall g$$ Latency

$$\mathbb{E}_g[\varepsilon_2(n_1(g), n_2(g), P_1(g), P_2(g), g)] \leq \varepsilon_{\mathrm{rel}}$$ Reliability

Optimization Problem, *Full-CSI*

Fixing: B information bits, M = 2, $D = 0$

$$\min_{n_1(g), n_2(g), P_1(g), P_2(g)} \mathbb{E}_g\Big[-\frac{\overbrace{\mathcal{T}_h(a)}^{\boxed{(1-a)\left(\frac{B(1-\varepsilon_2)}{n_1+n_2\varepsilon_1}\right)}}}{\mathcal{T}_{h,max}} + \frac{\overbrace{\mathcal{E}(a)}^{\boxed{a\,(n_1P_1+n_2P_2\varepsilon_1)}}}{\mathcal{E}_{min}}\Big]$$

$$\begin{aligned}
s.t.\quad & n_1(g) + n_2(g) \leq \mathrm{N}_\ell, \quad \forall g && \text{Latency}\\
& \mathbb{E}_g[\varepsilon_2(n_1(g), n_2(g), P_1(g), P_2(g), g)] \leq \varepsilon_{\mathrm{rel}} && \text{Reliability}\\
& n_1(g)P_1(g) + n_2(g)P_2(g) \leq \mathrm{E}_b, \quad \forall g && \text{Energy}\\
& P_i(g) \leq \mathrm{P}_{\max}, \quad i \in \{1,2\} \quad \forall g && \text{Maximum Power}
\end{aligned}$$

Optimization Problem, *Statistical-CSI*

Fixing: B information bits, M = 2, $D = 0$

$$\min_{n_1\cancel{(g)},n_2\cancel{(g)},P_1\cancel{(g)},P_2\cancel{(g)}} \mathbb{E}_g[-\frac{\mathcal{T}_h(a)}{\mathcal{T}_{h,max}} + \frac{\mathcal{E}(a)}{\mathcal{E}_{min}}]$$

$$\textit{s.t.} \quad n_1\cancel{(g)} + n_2\cancel{(g)} \leq \mathrm{N}_\ell, \quad \forall g \quad \text{Latency}$$

$$\mathbb{E}_g[\varepsilon_2(n_1\cancel{(g)}, n_2\cancel{(g)}, P_1\cancel{(g)}, P_2\cancel{(g)}, g)] \leq \varepsilon_{\mathrm{rel}} \quad \text{Reliability}$$

$$n_1\cancel{(g)} P_1\cancel{(g)} + n_2\cancel{(g)} P_2\cancel{(g)} \leq \mathrm{E}_b, \quad \forall g \quad \text{Energy}$$

$$P_i\cancel{(g)} \leq \mathrm{P}_{\max}, \quad i \in \{1,2\} \quad \forall g \quad \text{Maximum Power}$$

The full-CSI problem is an optimization problem with *functions as solution*:

Simplifying Assumption:

The full-CSI problem is an optimization problem with *functions as solution*:

Simplifying Assumption:

Reliability Constraint: $\mathbb{P}(g < g_{th}) + \mathbb{P}(g \geq g_{th})\varepsilon_{on} \leq \varepsilon_{\text{rel}}$

Optimization Variables: $g_{th}, (n_1, n_2, P_1, P_2) \;\; \forall g \geq g_{th}$

Lemma to find feasibility set

Lemma: *The solution of the problem:*

$$\min_{n_1,\ldots,n_M,P_1,\ldots,P_M,M} \quad \varepsilon_M(n_1,\ldots n_m,P_1,\ldots P_m,g)$$

$$\text{s.t.} \quad \sum_{i=1}^{M} n_i \leq \mathrm{N}_\ell$$

$$\sum_{i=1}^{M} n_i P_i \leq \mathrm{E}_b$$

is $M = 1$ *with* $(n_1, P_1) = (\mathrm{N}_\ell, \frac{\mathrm{E}_b}{\mathrm{N}_\ell})$.

Given hard* *energy and latency constraint, using all the resources on one packet minimizes the error.*

*In [3]*A. Avranas, M. Kountouris, and P. Ciblat, "Energy-latency tradeoff in ultra-reliable low-latency communication with retransmissions," IEEE JSAC., Oct. 2018 we used soft constraints for a similar problem leading to quite different behavior.*

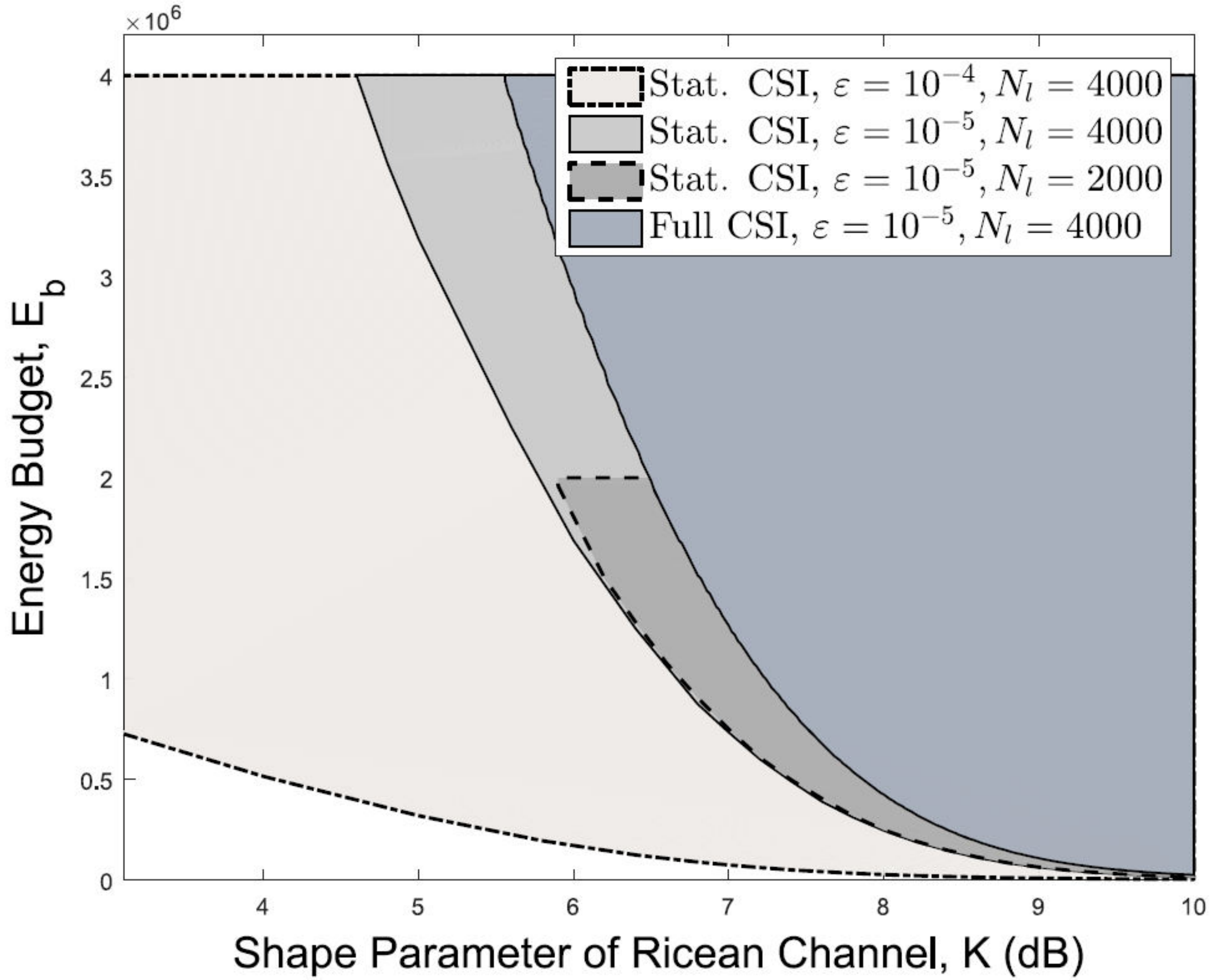


Fig. 1: Feasibility region for different channel, $B = 32$Bytes, maximum energy budget $\mathrm{E}_b = P_{\max}\mathrm{N}_\ell$ with $P_{\max} = 30dB$.

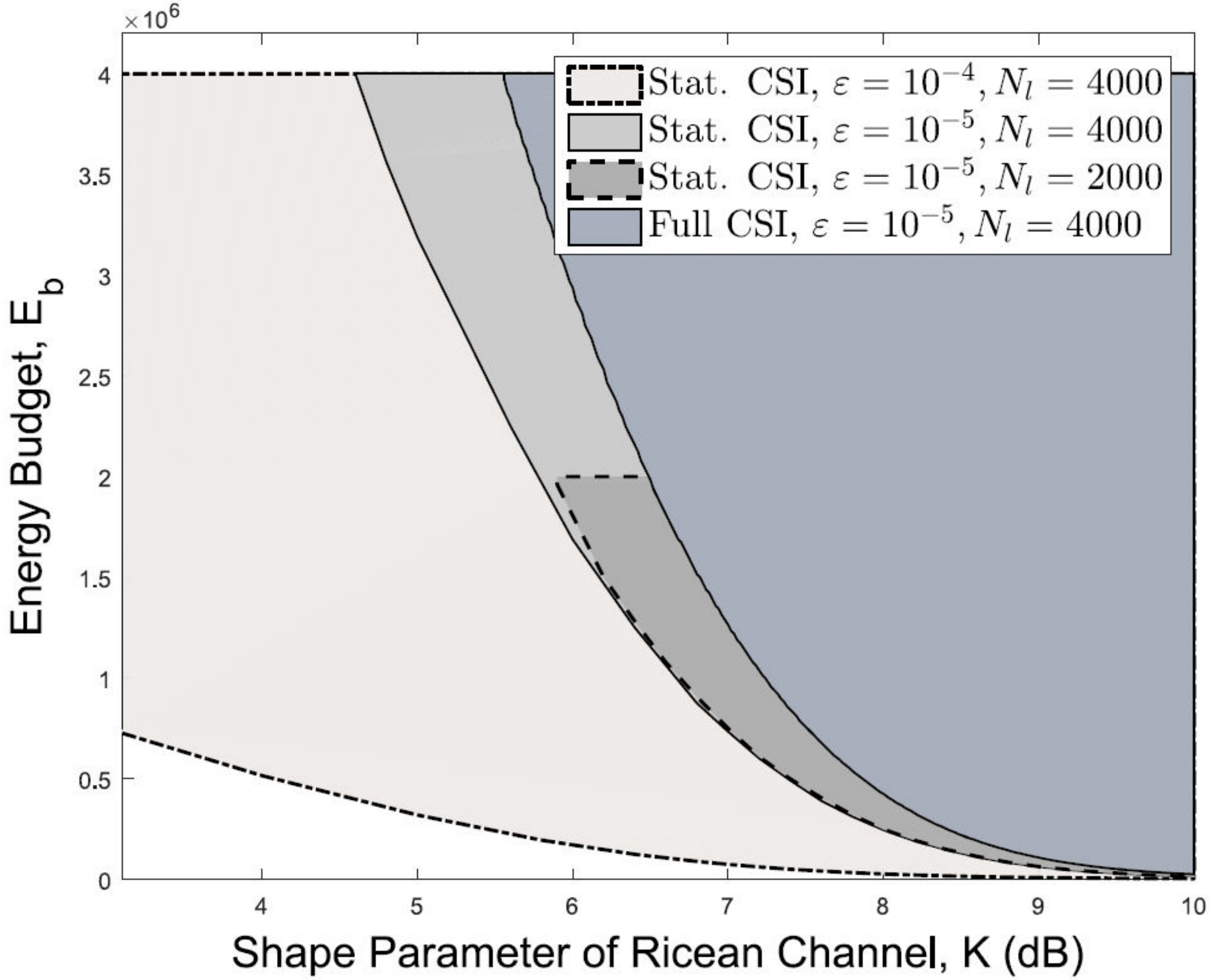


Fig. 1: Feasibility region for different channel, $B = 32$Bytes, maximum energy budget $\mathrm{E}_b = P_{\max}\mathrm{N}_\ell$ with $P_{\max} = 30dB$.

The trade-off and Pareto frontiers

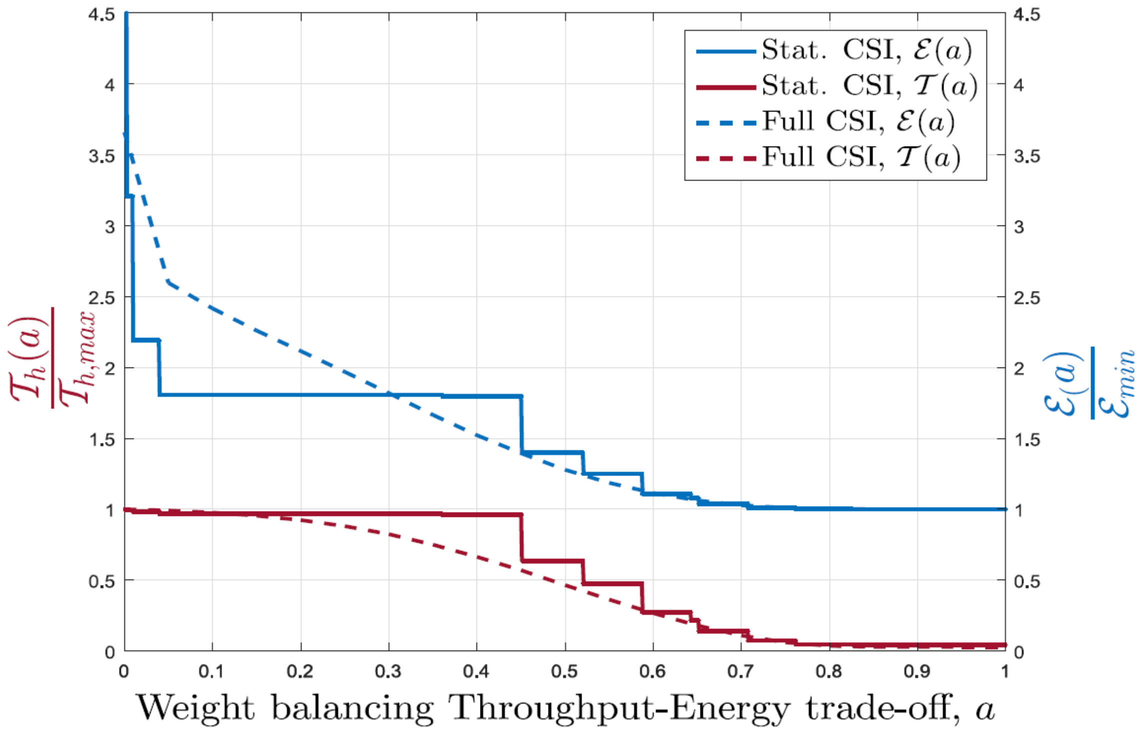


Fig. 2: Throughput and energy relative to their optimal value for Rician channel with $K = 7$dB, $B = 32$Bytes, $\varepsilon_{\text{rel}} = 10^{-5}$ and maximum energy $\text{E}_b = P_{\max}\text{N}_\ell$ with $P_{\max} = 30dB$ and $\text{N}_\ell = 4000$.

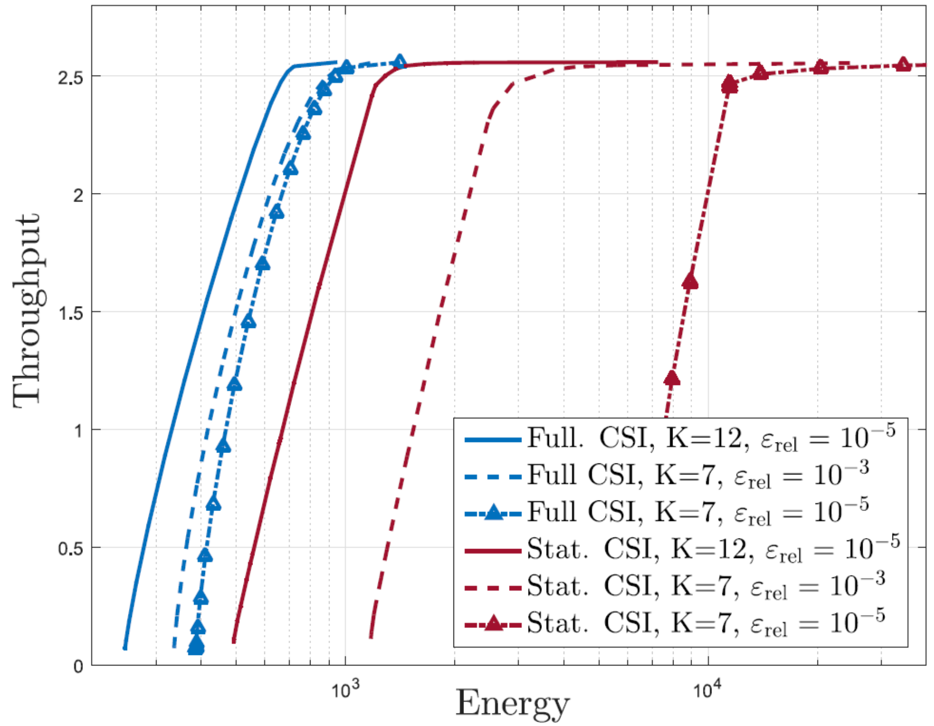


Fig. 3: Pareto frontier for throughput and energy, with $\text{E}_b = P_{\max}\text{N}_\ell$, $P_{\max} = 30dB$, and $\text{N}_\ell = 4000$.

IR-HARQ vs One-Shot

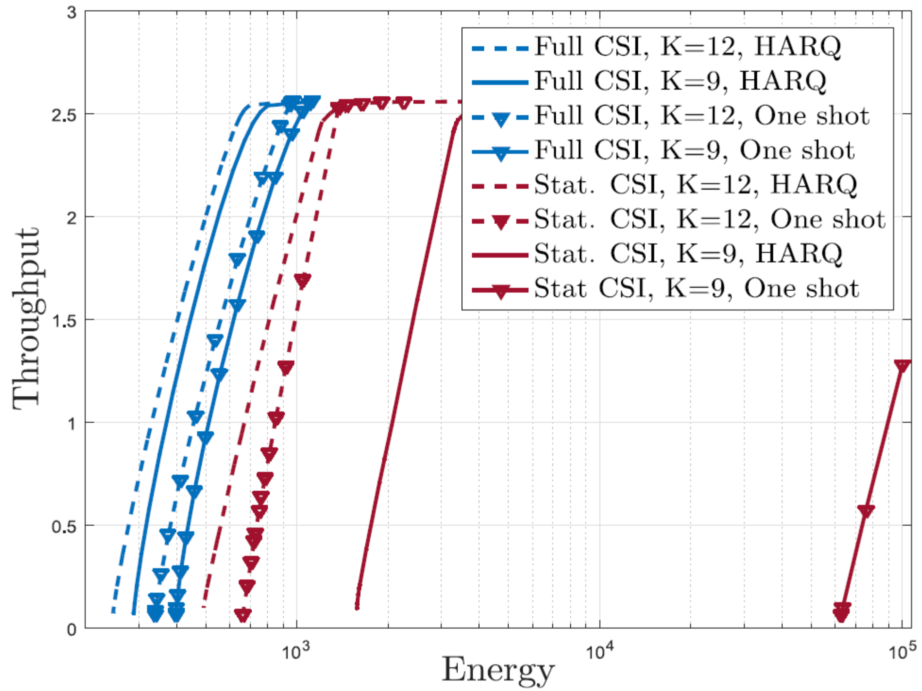


Fig. 4: Pareto frontier for throughput and energy when HARQ or one shot transmission is used, with $\mathrm{E}_b = P_{\max}\mathrm{N}_\ell$, $P_{\max} = 1000(30dB)$, and $\mathrm{N}_\ell = 4000$.

Thank you

www.huawei.com