GreenWeb:

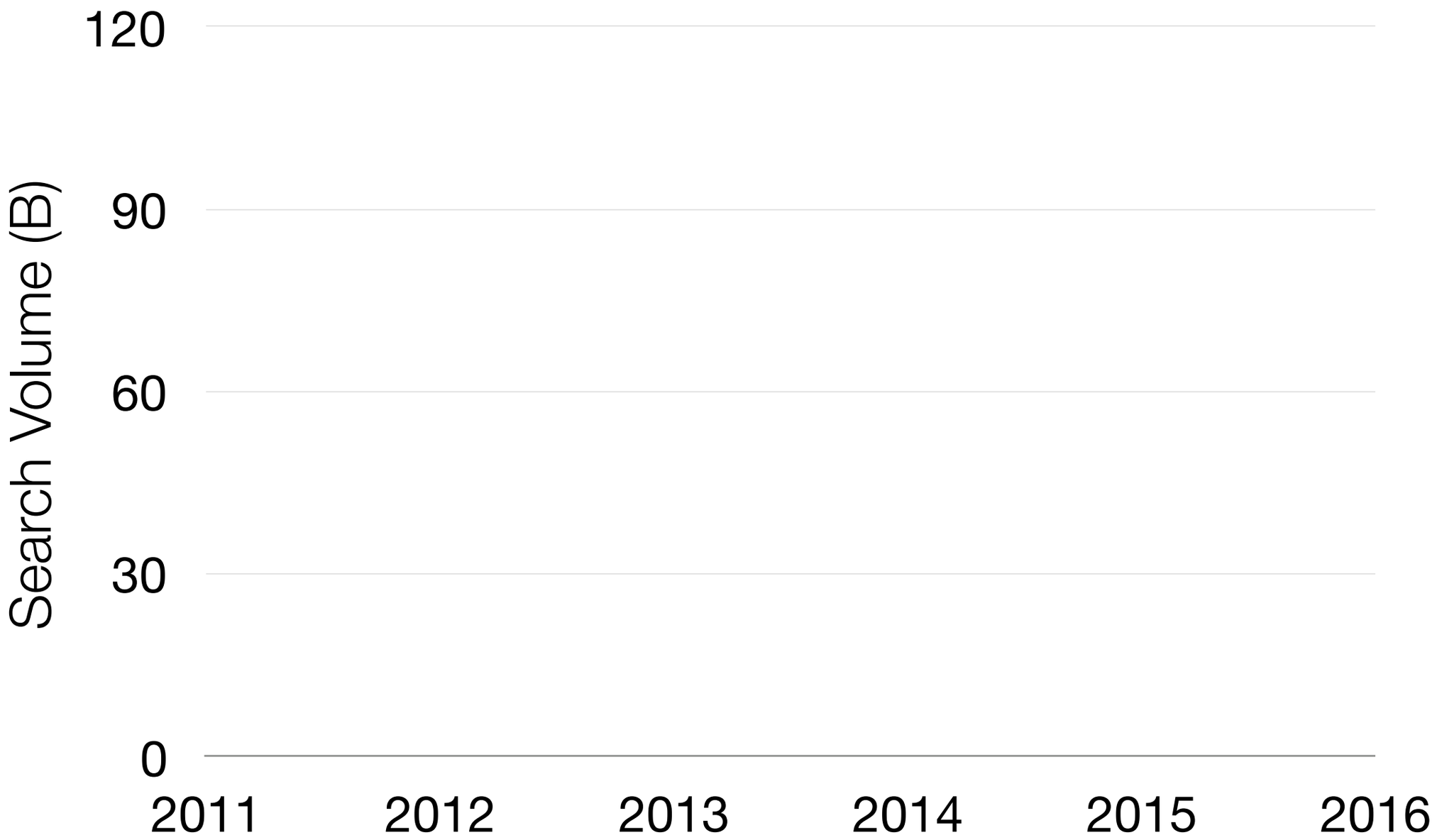
Language Extensions for Energy-Efficient Mobile Web Computing

Yuhao Zhu
The University of Texas at Austin
with Vijay Janapa Reddi

PLDI 2016

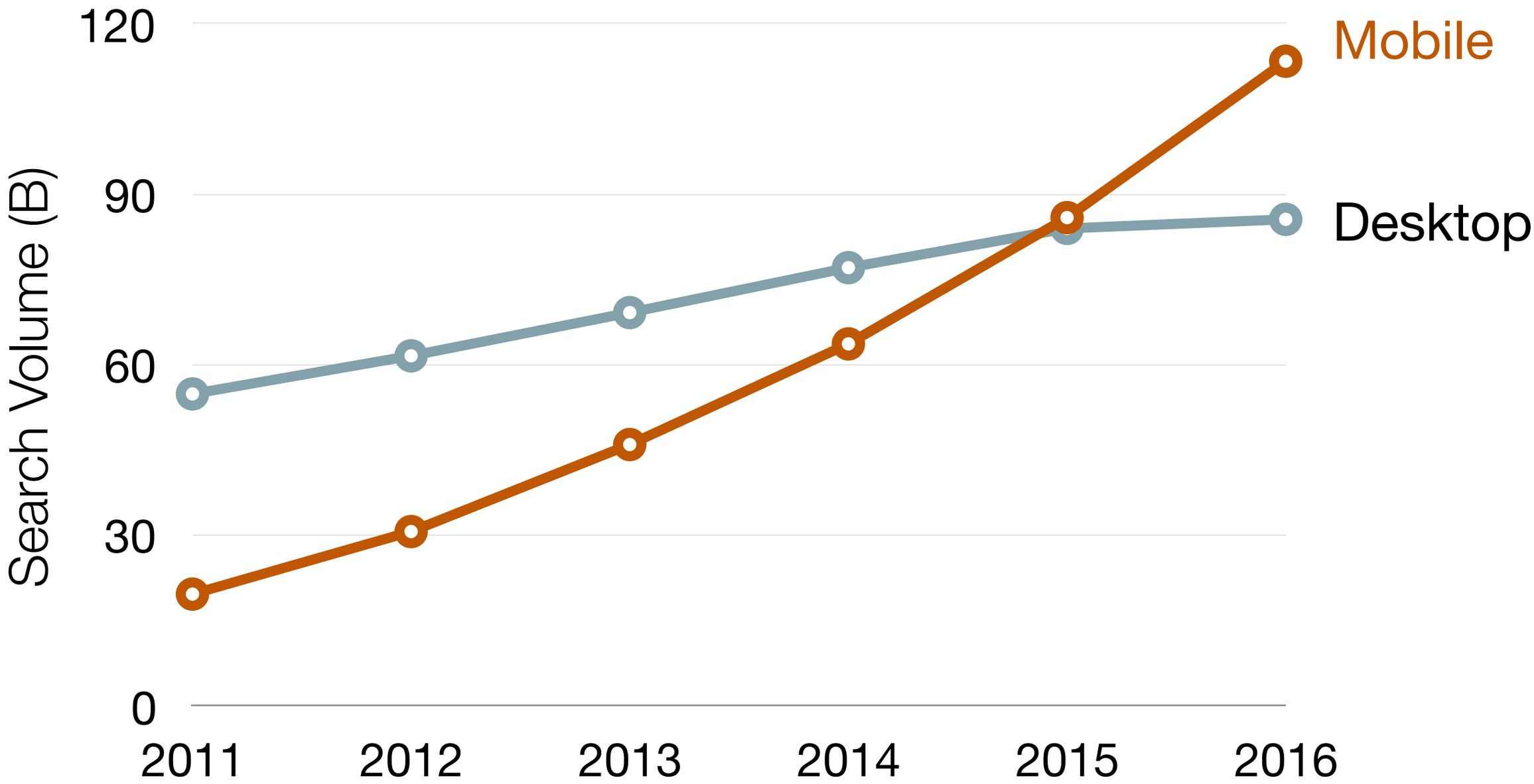
Web: Mobile Overtaking Desktop

Web: Mobile Overtaking Desktop



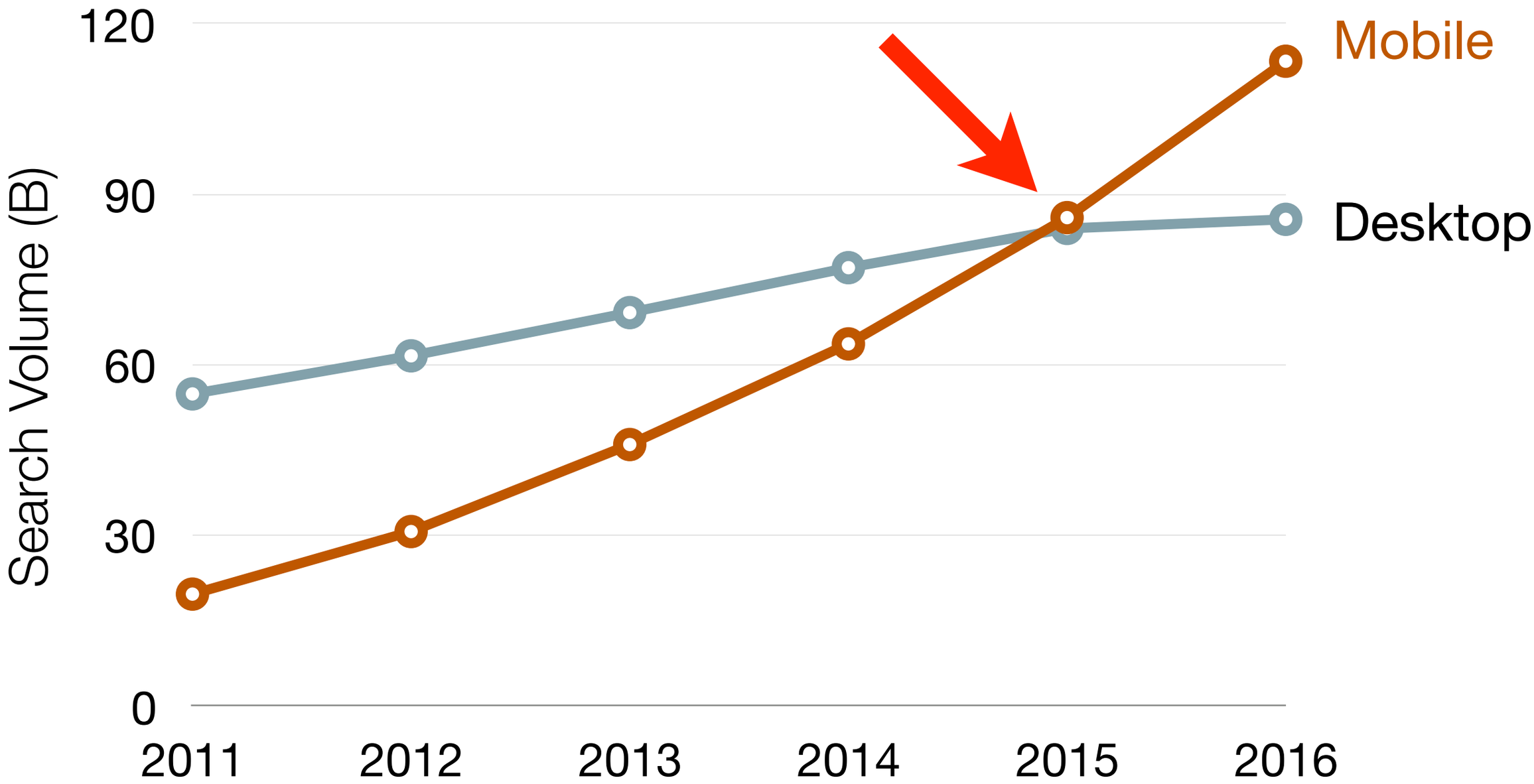
Source: BIA/Kelsey

Web: Mobile Overtaking Desktop



Source: BIA/Kelsey

Web: Mobile Overtaking Desktop



Source: BIA/Kelsey

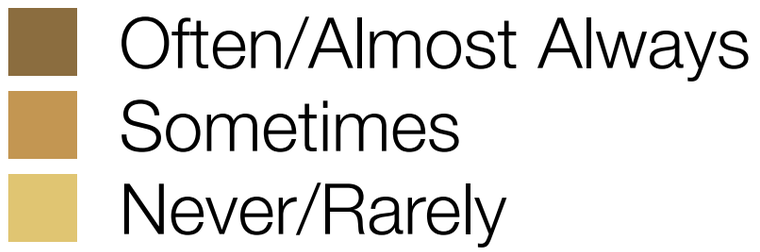
Web ≈ Mobile Web

Energy Concern Among Mobile Developers

[ICSE 2016] Manotas et al., “An Empirical Study of Practitioners’ Perspectives on Green Software Engineering”

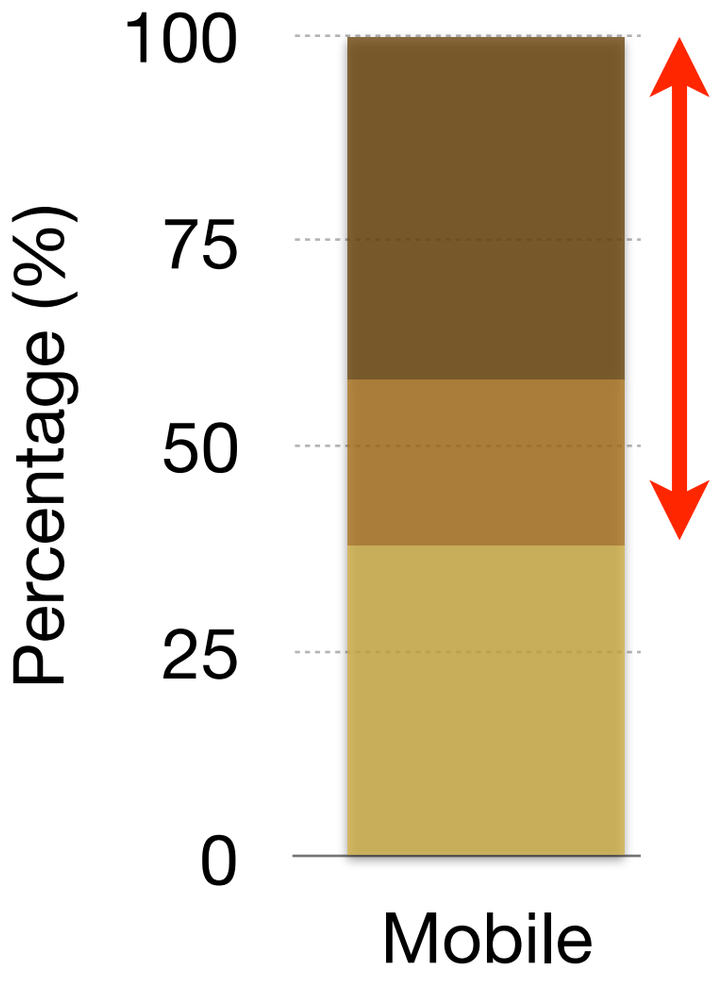
Energy Concern Among Mobile Developers

"My applications have requirements about energy usage."



[ICSE 2016] Manotas et al., "An Empirical Study of Practitioners' Perspectives on Green Software Engineering"

Energy Concern Among Mobile Developers

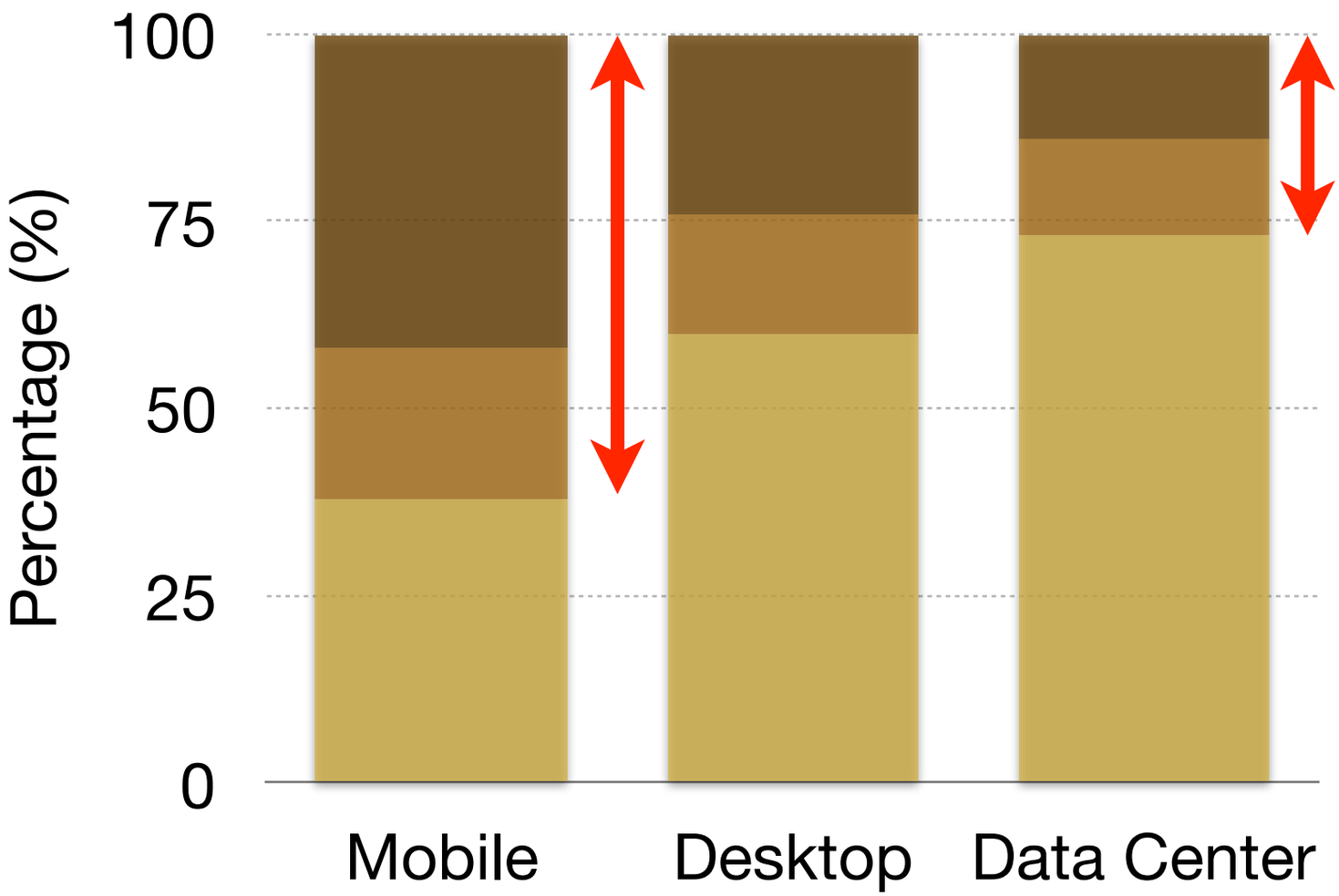


"My applications have requirements about energy usage."

- Often/Almost Always
- Sometimes
- Never/Rarely

[ICSE 2016] Manotas et al., "An Empirical Study of Practitioners' Perspectives on Green Software Engineering"

Energy Concern Among Mobile Developers



"My applications have requirements about energy usage."

- Often/Almost Always
- Sometimes
- Never/Rarely

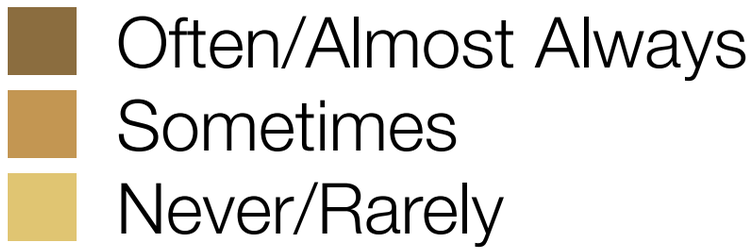
[ICSE 2016] Manotas et al., "An Empirical Study of Practitioners' Perspectives on Green Software Engineering"

Developers are Willing to Make Trade-offs

[ICSE 2016] Manotas et al., “An Empirical Study of Practitioners’ Perspectives on Green Software Engineering”

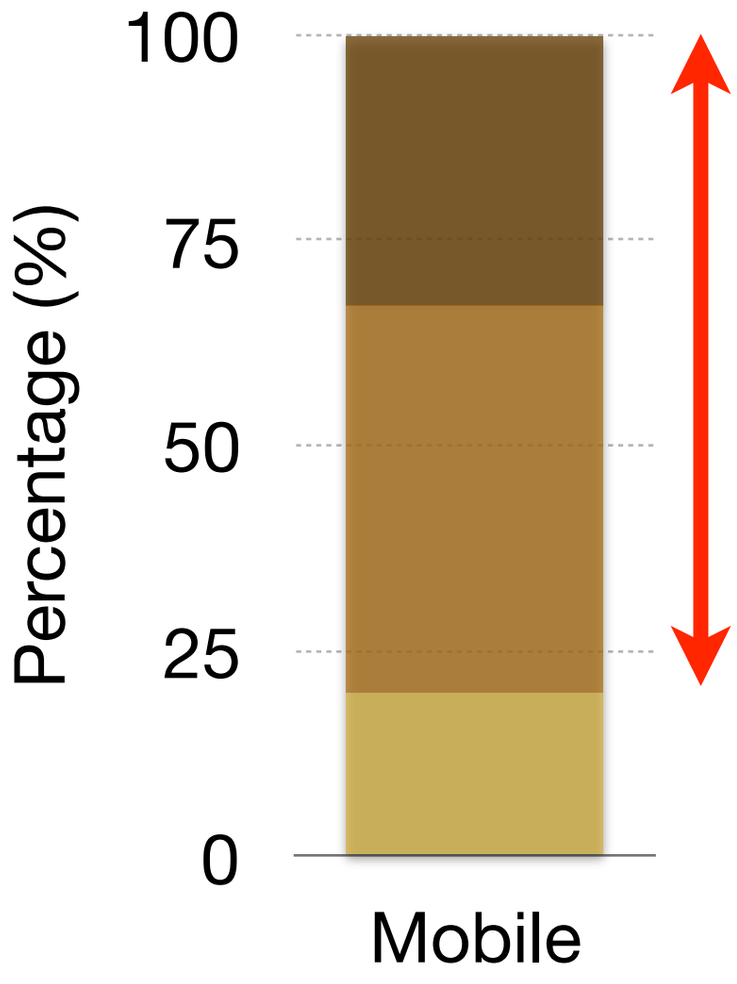
Developers are Willing to Make Trade-offs

"I'm willing to sacrifice performance, etc. for reduced energy usage."



[ICSE 2016] Manotas et al., "An Empirical Study of Practitioners' Perspectives on Green Software Engineering"

Developers are Willing to Make Trade-offs



"I'm willing to sacrifice performance, etc. for reduced energy usage."

Often/Almost Always
Sometimes
Never/Rarely

[ICSE 2016] Manotas et al., "An Empirical Study of Practitioners' Perspectives on Green Software Engineering"

Energy-efficiency

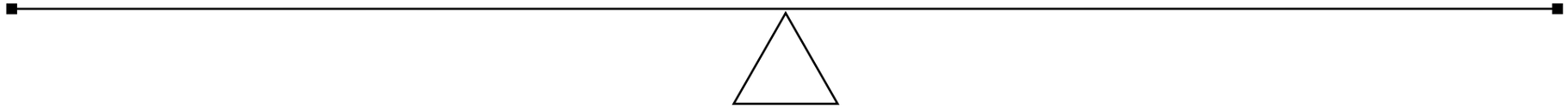
Quality-of-service

Energy-efficiency

Quality-of-service

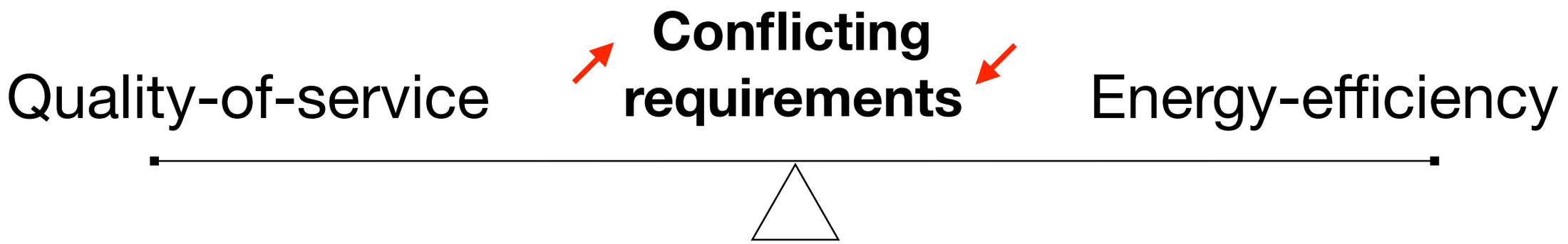
Conflicting requirements

Energy-efficiency



GreenWeb

Programming language support for balancing energy-efficiency and QoS in mobile Web computing



GreenWeb

Programming language support for balancing energy-efficiency and QoS in mobile Web computing

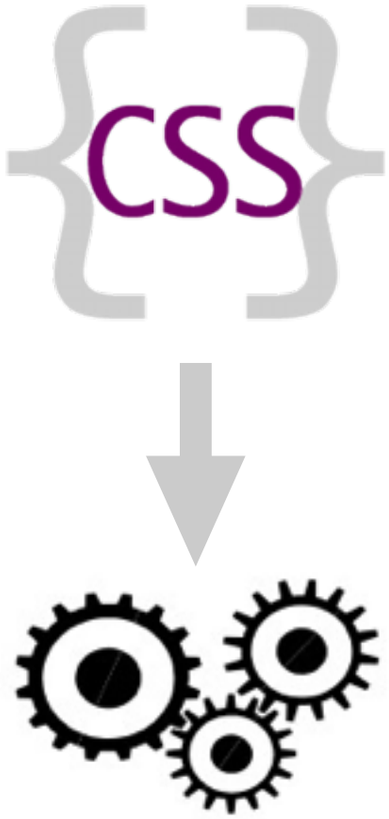
GreenWeb

Programming language support for balancing energy-efficiency and QoS in mobile Web computing

GreenWeb: Language for Energy-Efficiency

- **Language abstractions** for expressing QoS

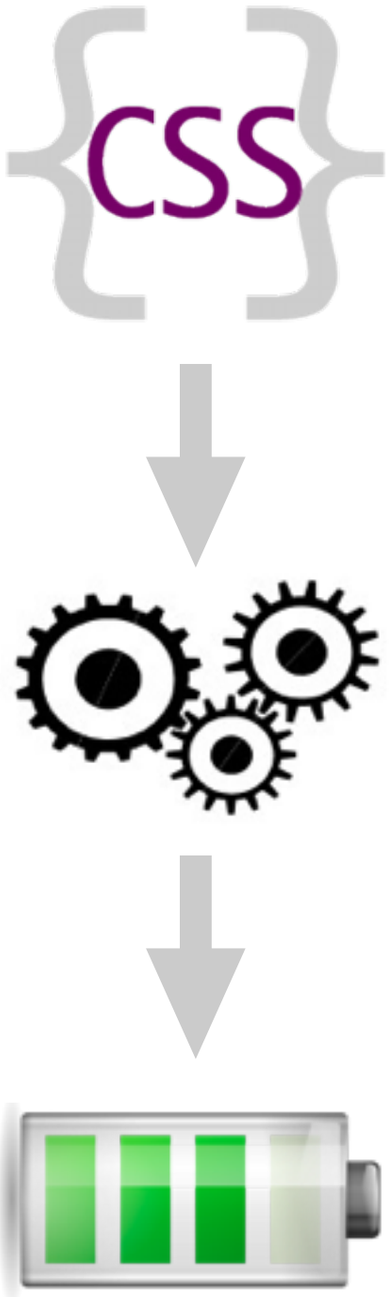
GreenWeb: Language for Energy-Efficiency



- **Language abstractions** for expressing QoS

- **Runtime** that saves energy while meeting the QoS constraints

GreenWeb: Language for Energy-Efficiency



- **Language abstractions** for expressing QoS
- **Runtime** that saves energy while meeting the QoS constraints
- **Result** in 60% energy savings on real hardware/software implementations

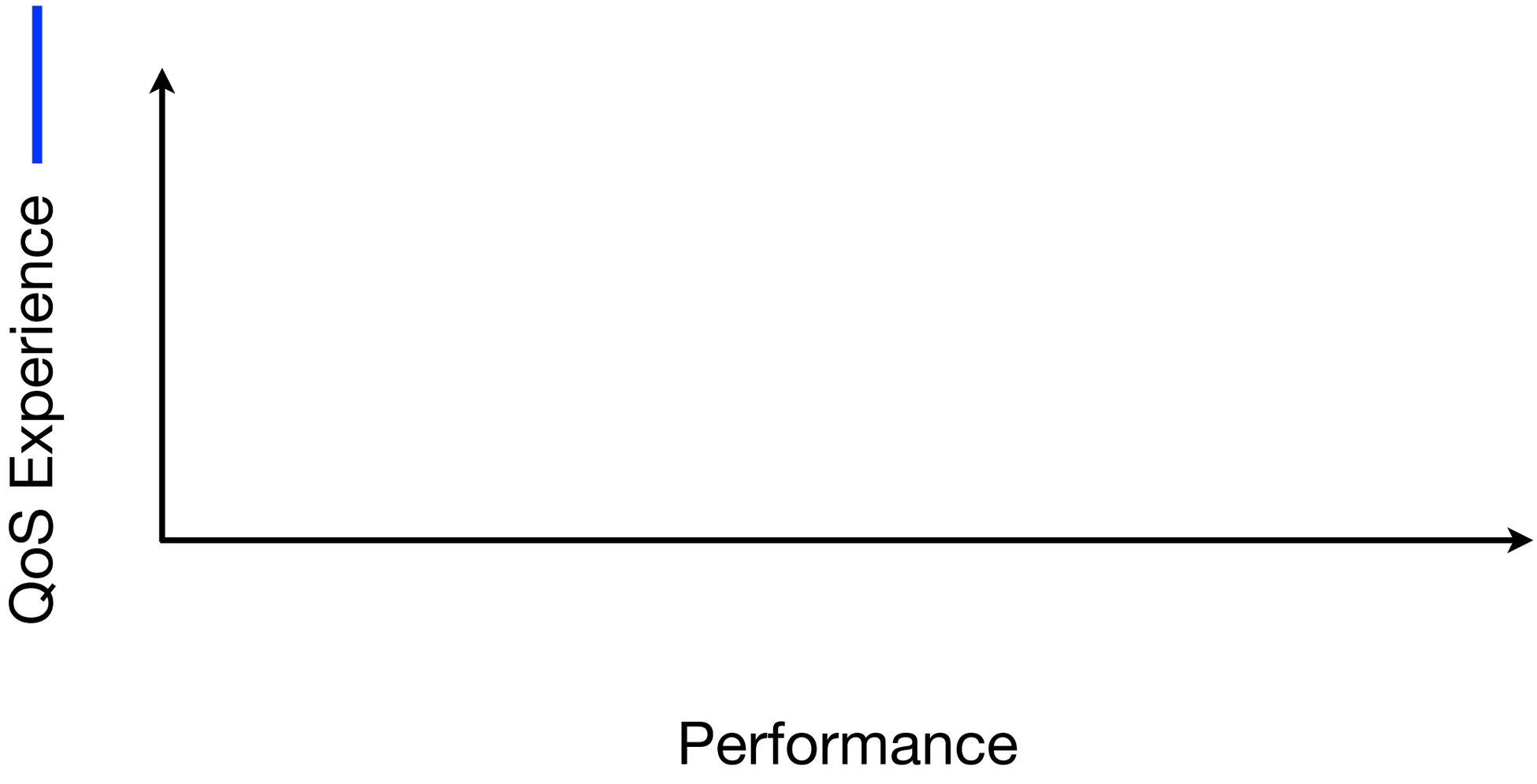
GreenWeb: Language for Energy-Efficiency

- **Language abstractions** for expressing QoS
- Runtime
  the QoS constraints
- Result
  hardware/software implementations

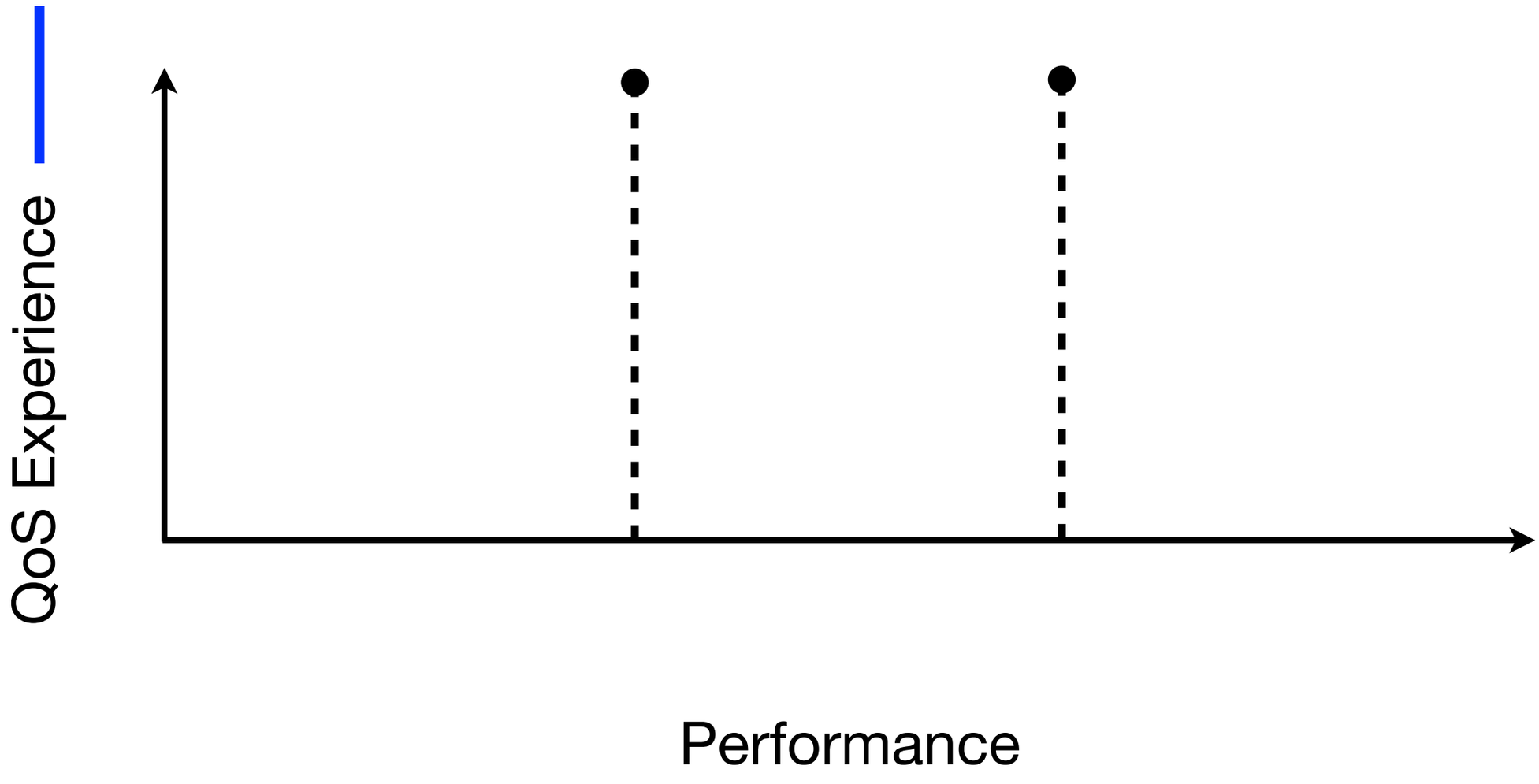
What is QoS in mobile Web?

Understanding Mobile Web QoS

Understanding Mobile Web QoS

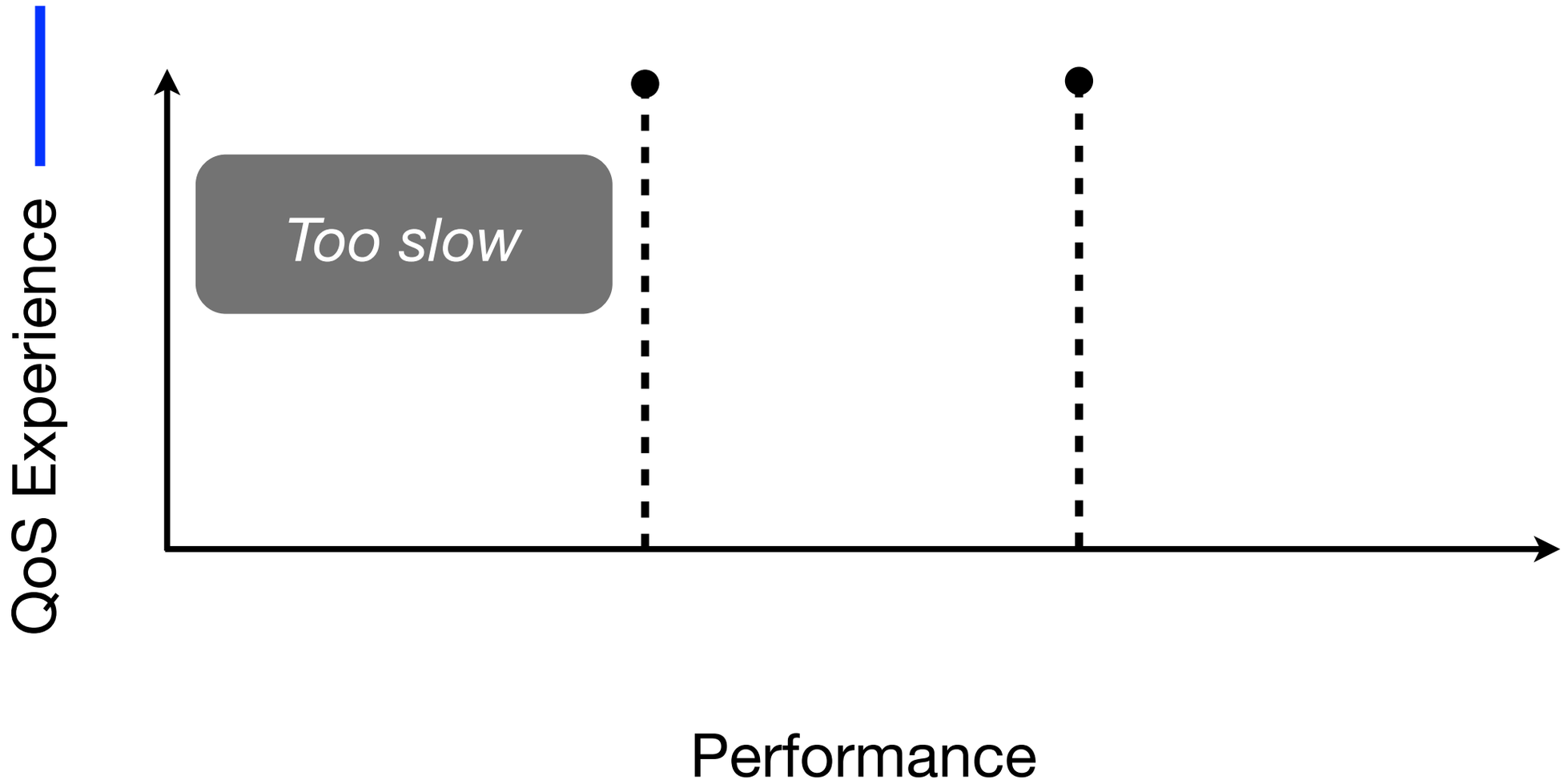


Understanding Mobile Web QoS



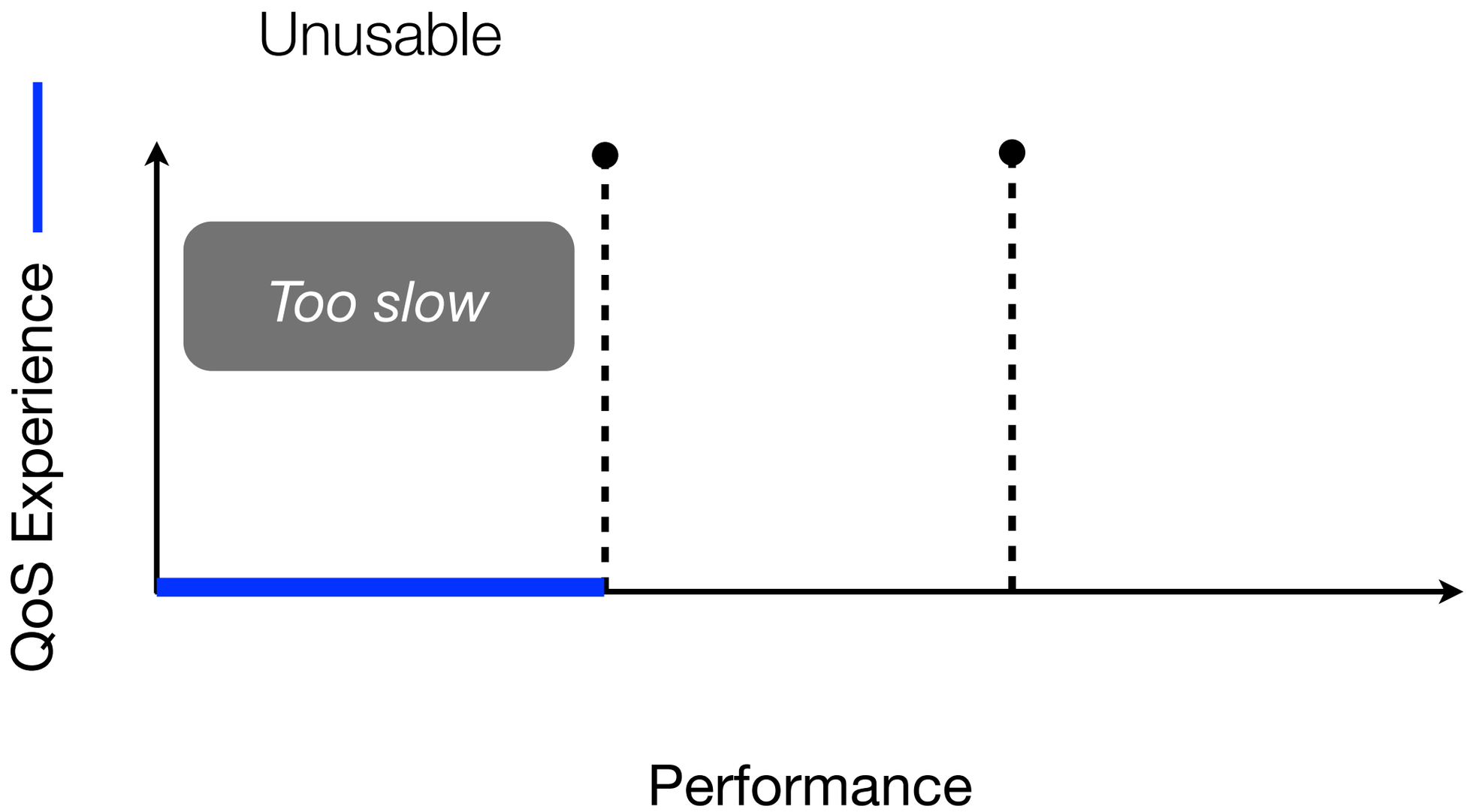
[OSDI 1996] Y. Endo et al., “Using Latency to Evaluate Interactive System Performance.”

Understanding Mobile Web QoS



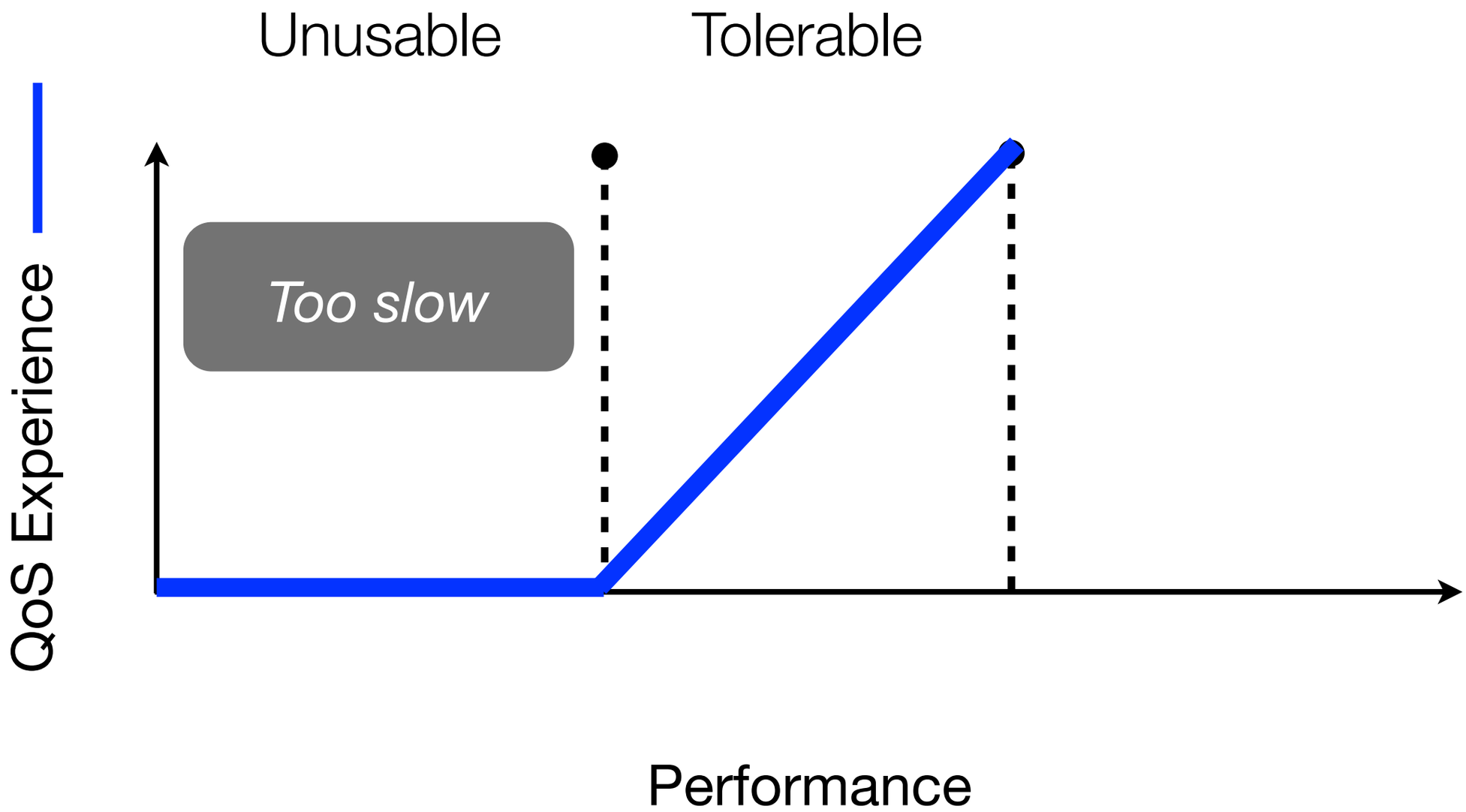
[OSDI 1996] Y. Endo et al., “Using Latency to Evaluate Interactive System Performance.”

Understanding Mobile Web QoS



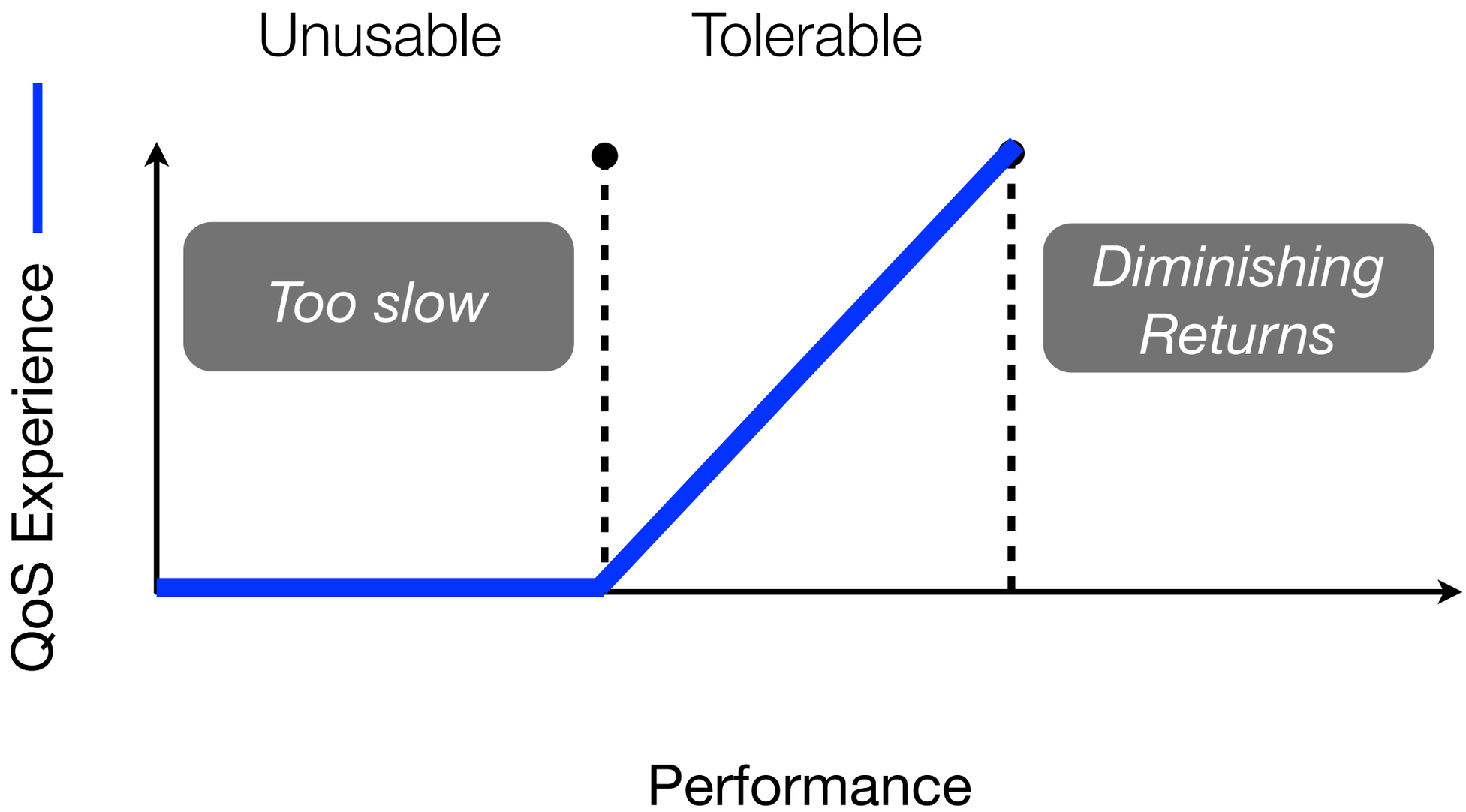
[OSDI 1996] Y. Endo et al., “Using Latency to Evaluate Interactive System Performance.”

Understanding Mobile Web QoS



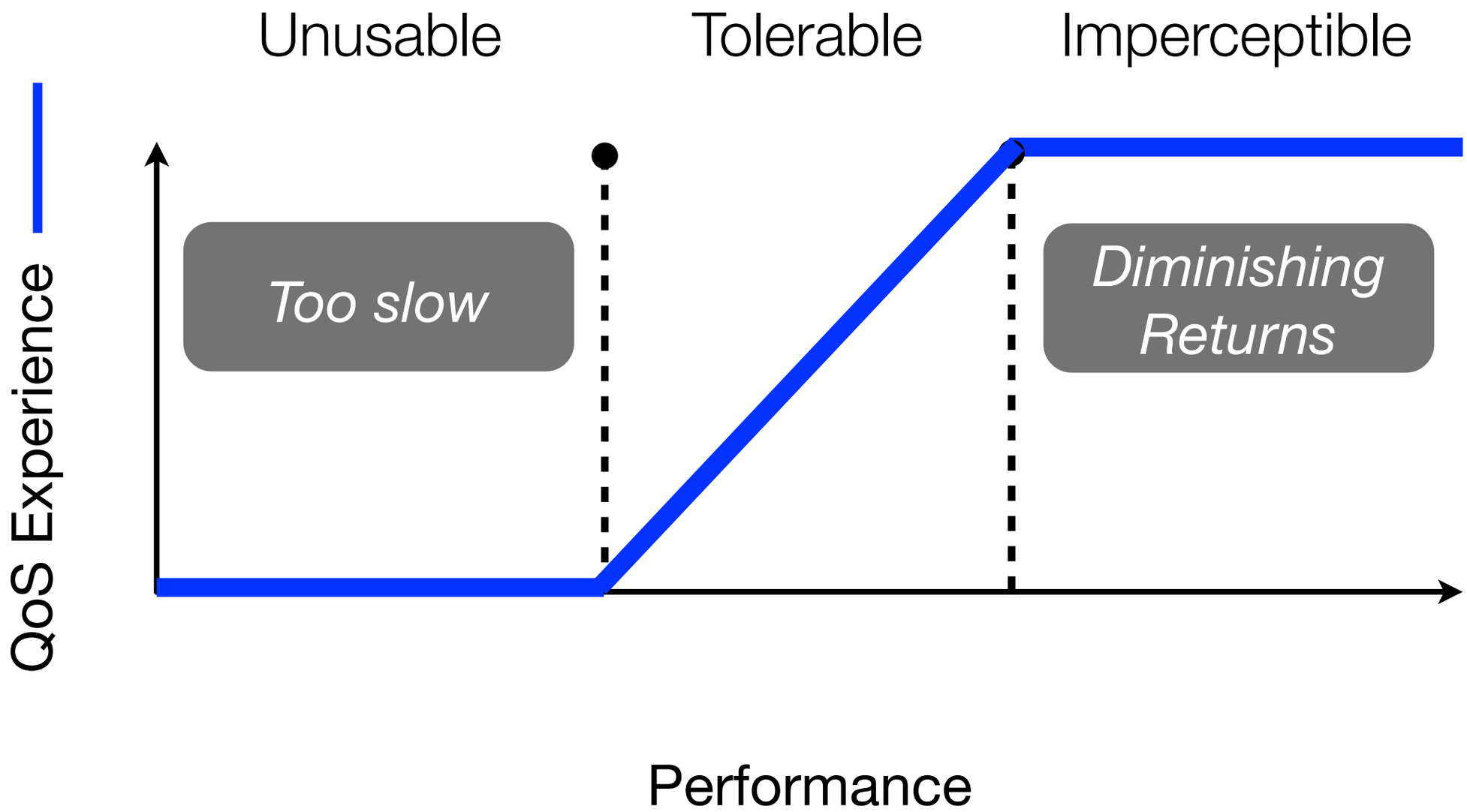
[OSDI 1996] Y. Endo et al., “Using Latency to Evaluate Interactive System Performance.”

Understanding Mobile Web QoS



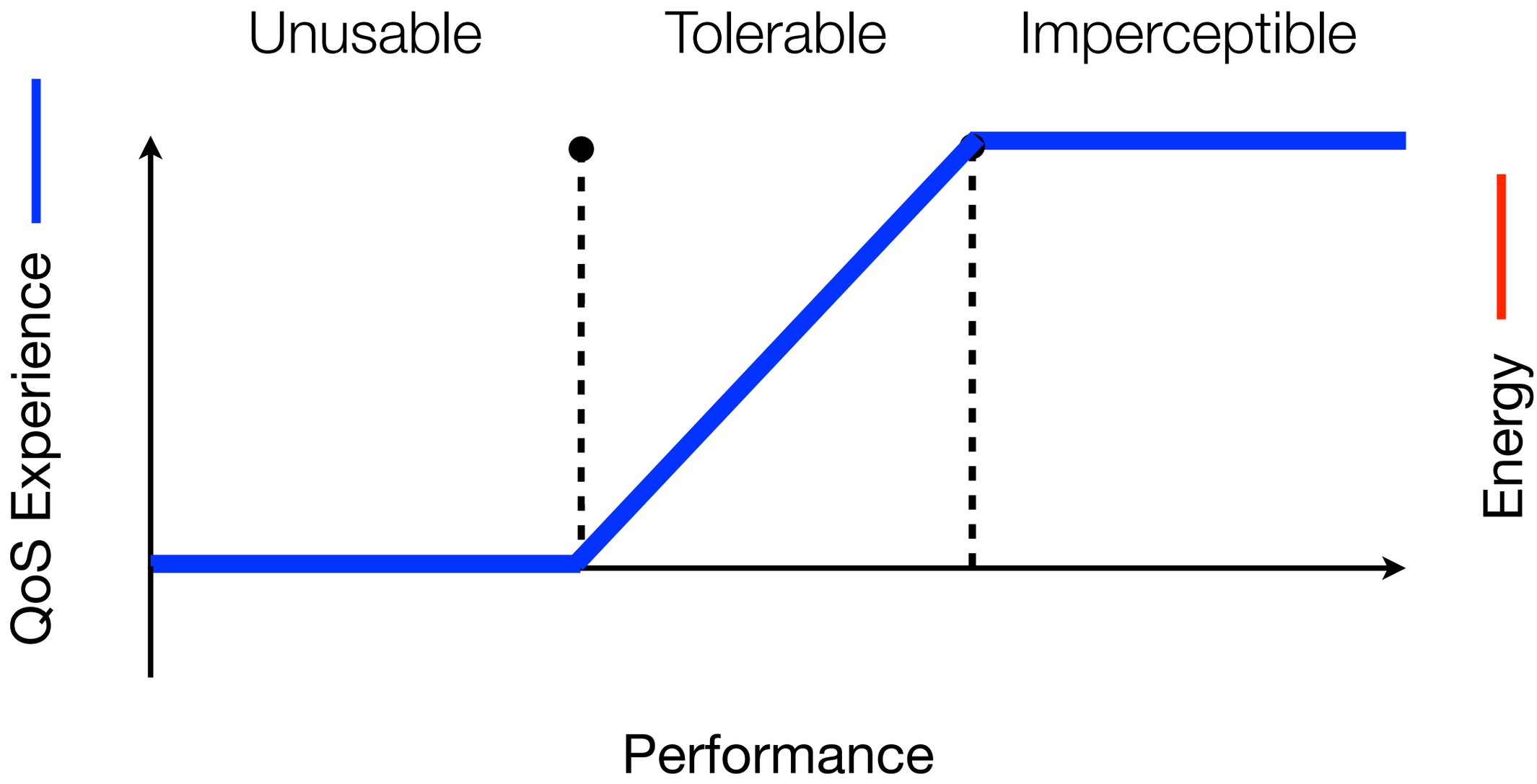
[OSDI 1996] Y. Endo et al., “Using Latency to Evaluate Interactive System Performance.”

Understanding Mobile Web QoS



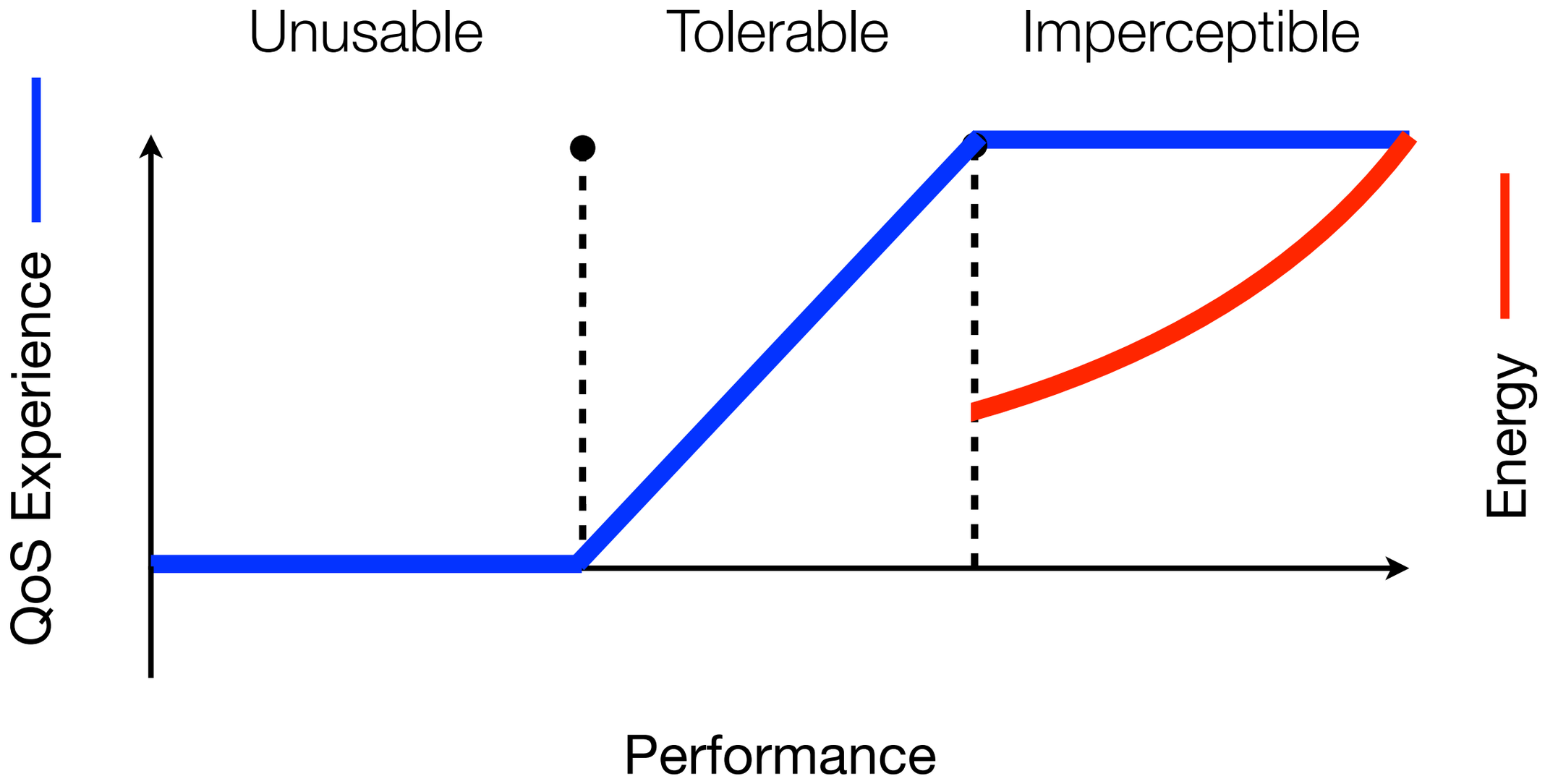
[OSDI 1996] Y. Endo et al., “Using Latency to Evaluate Interactive System Performance.”

Understanding Mobile Web QoS



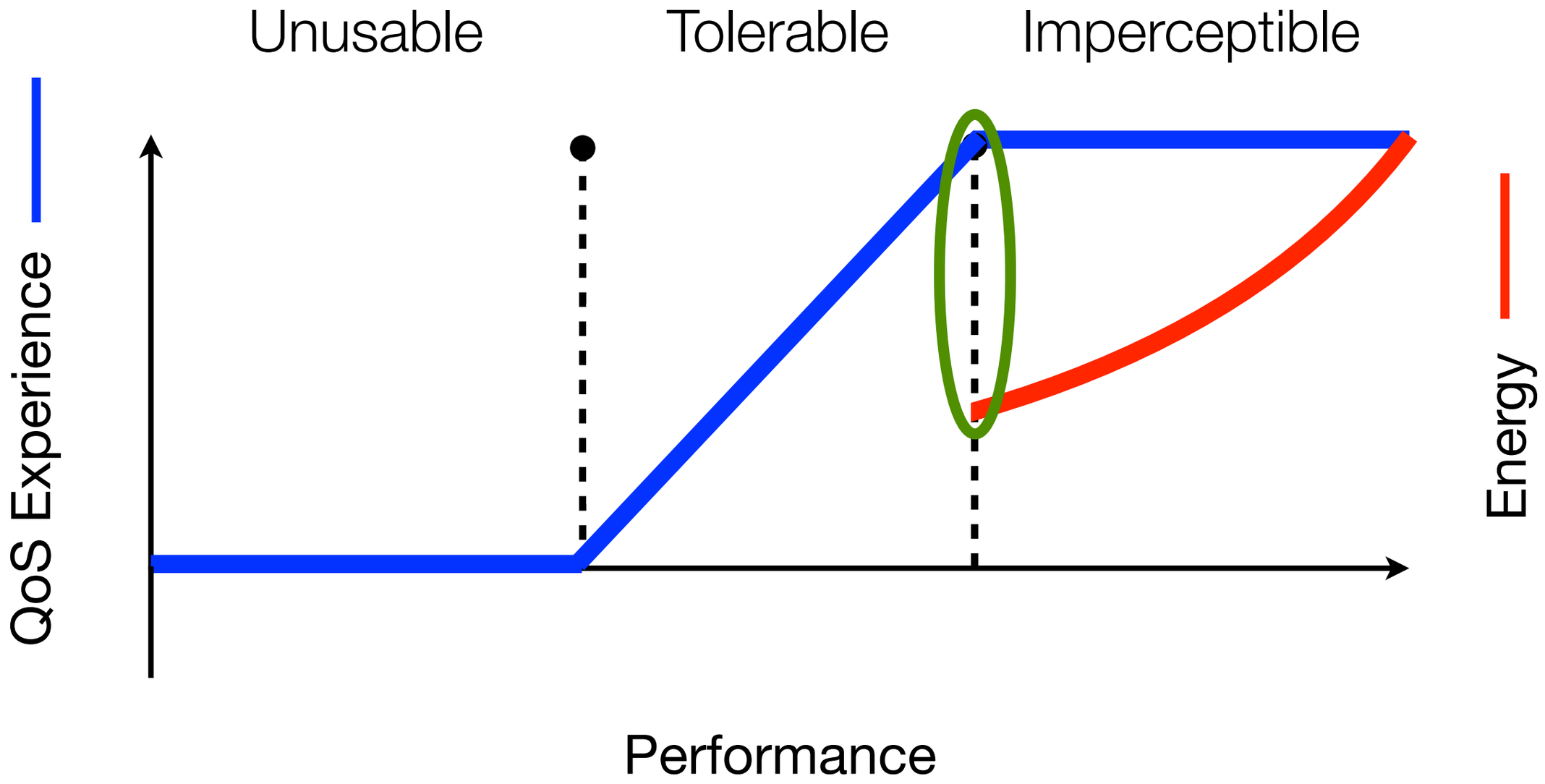
[OSDI 1996] Y. Endo et al., “Using Latency to Evaluate Interactive System Performance.”

Understanding Mobile Web QoS



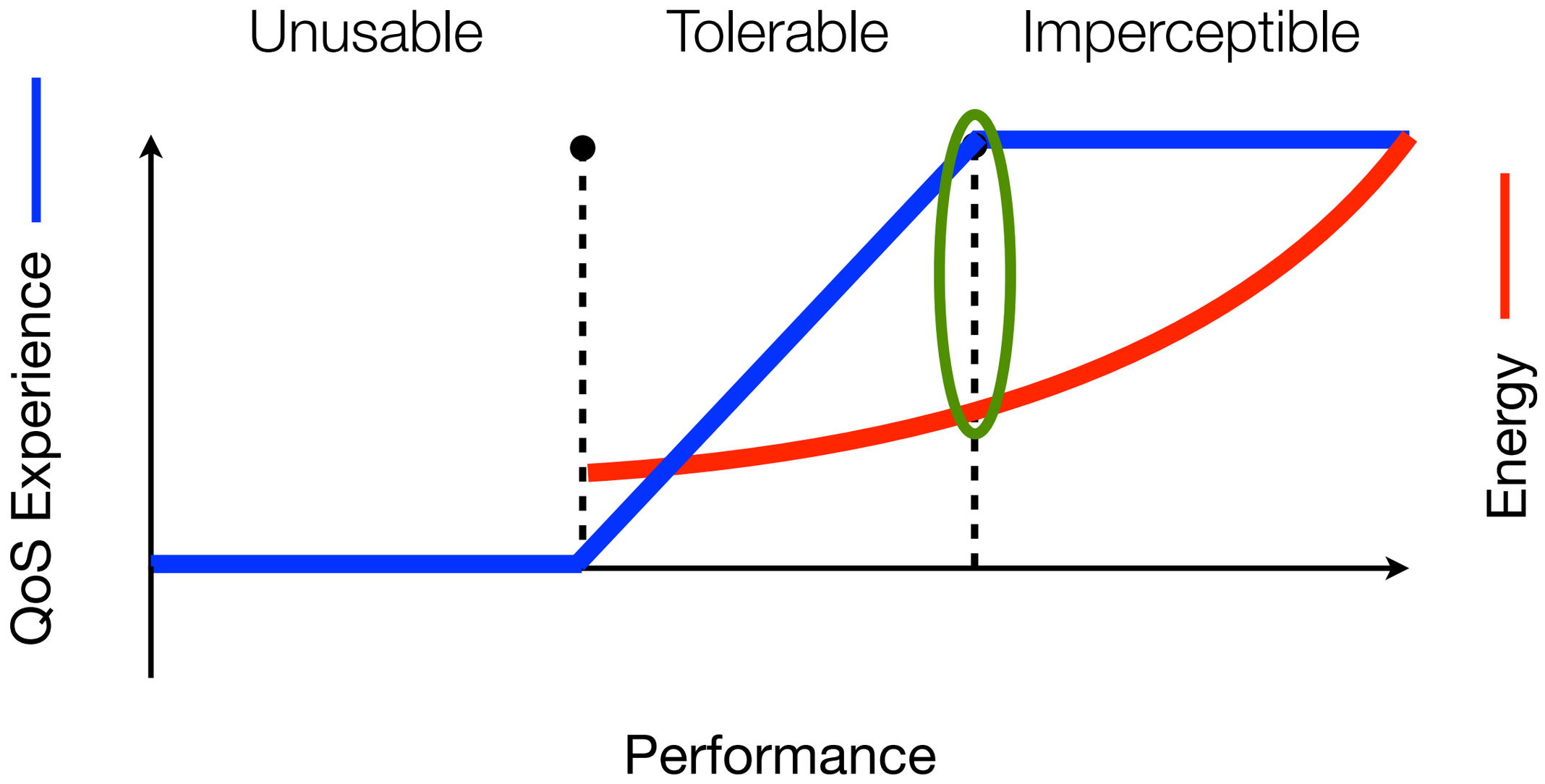
[OSDI 1996] Y. Endo et al., “Using Latency to Evaluate Interactive System Performance.”

Understanding Mobile Web QoS



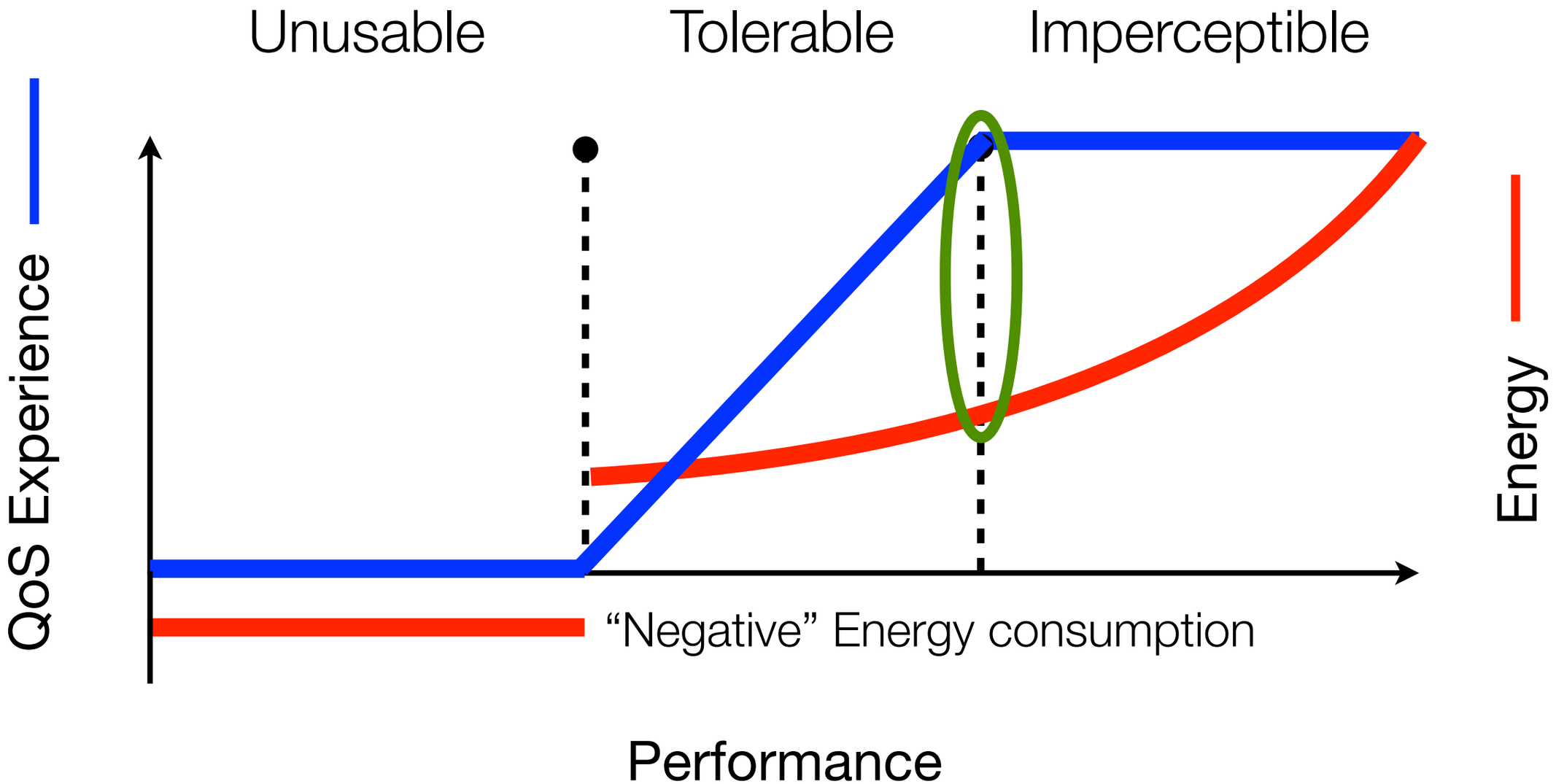
[OSDI 1996] Y. Endo et al., “Using Latency to Evaluate Interactive System Performance.”

Understanding Mobile Web QoS



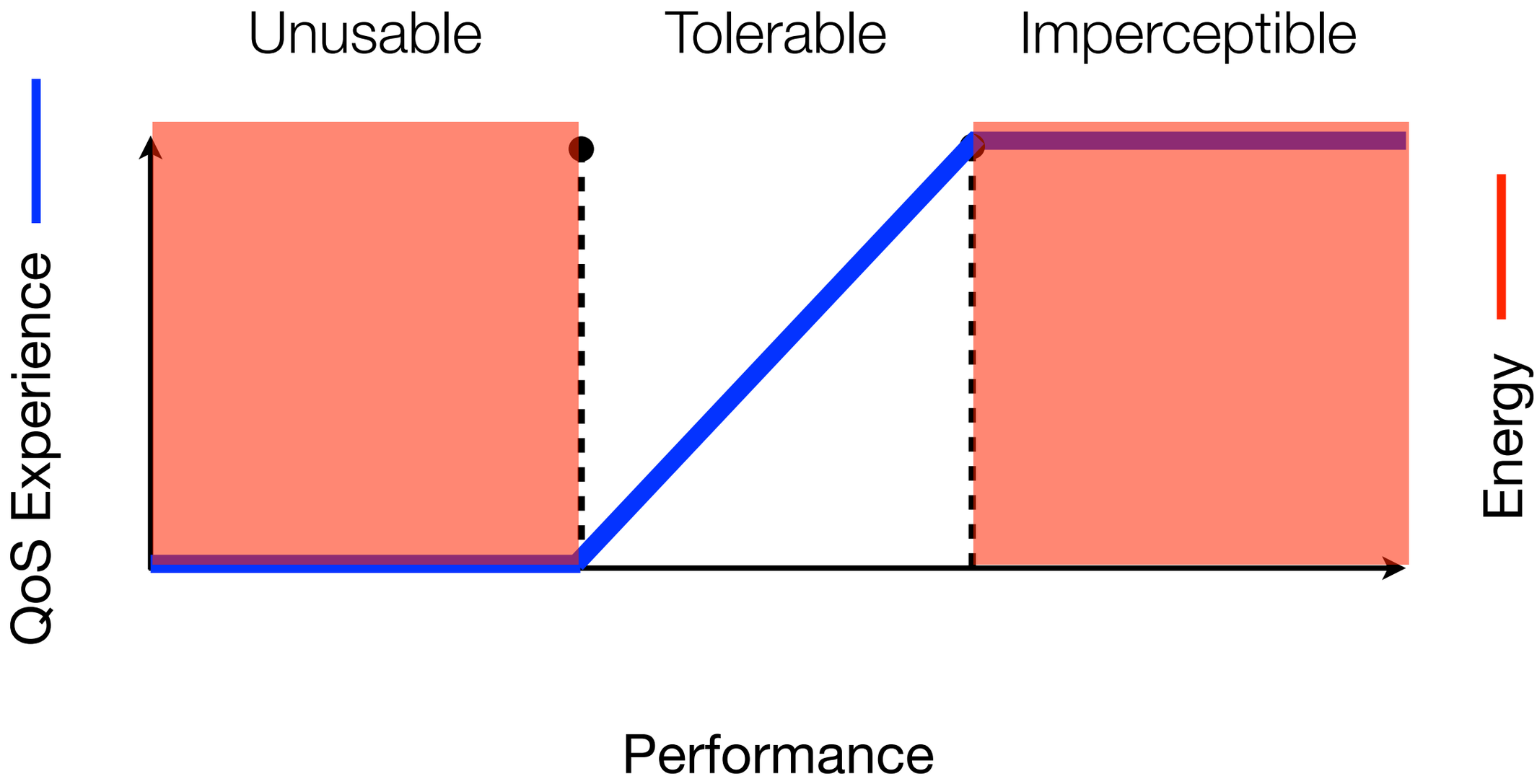
[OSDI 1996] Y. Endo et al., “Using Latency to Evaluate Interactive System Performance.”

Understanding Mobile Web QoS



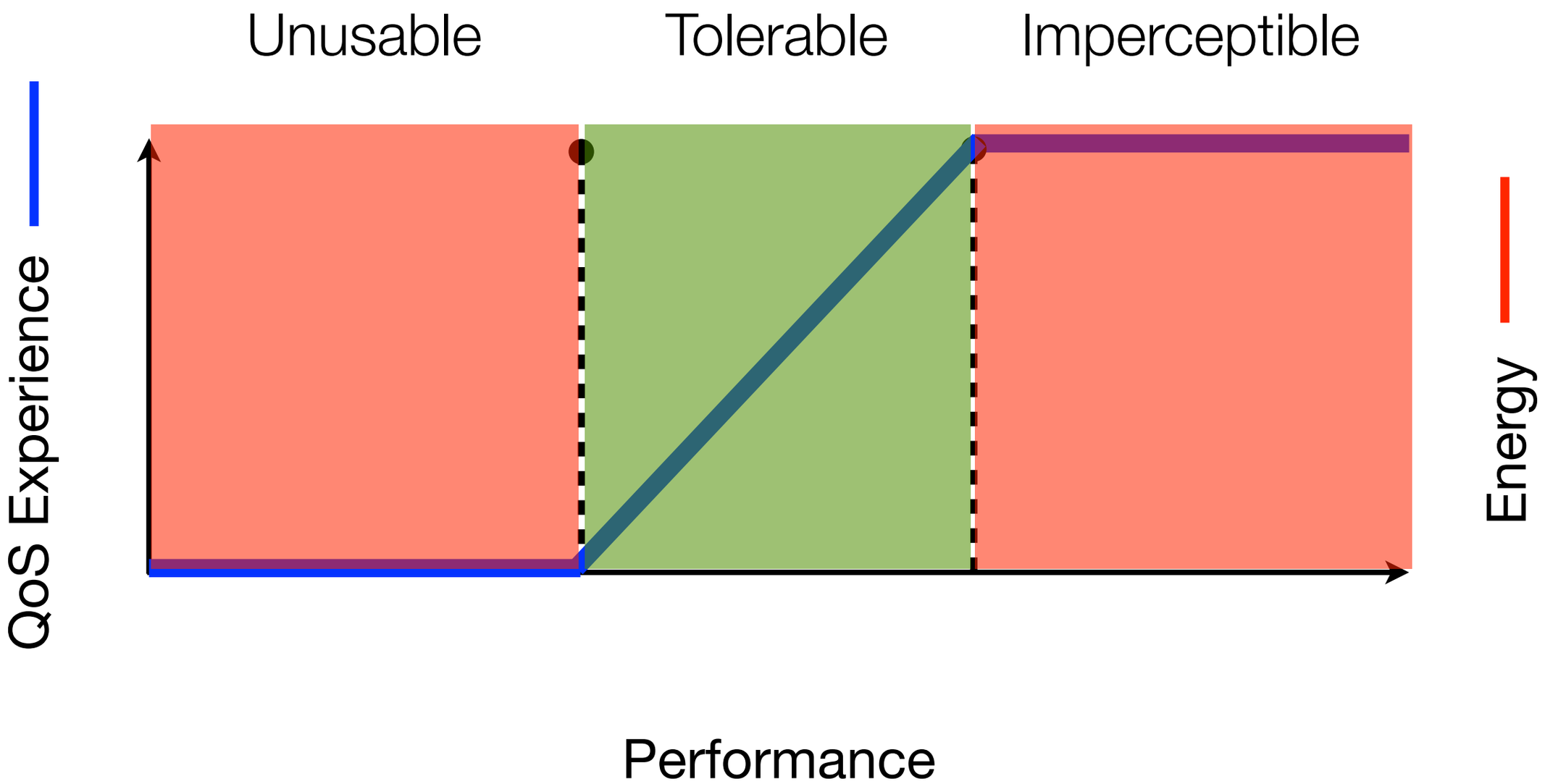
[OSDI 1996] Y. Endo et al., "Using Latency to Evaluate Interactive System Performance."

Understanding Mobile Web QoS



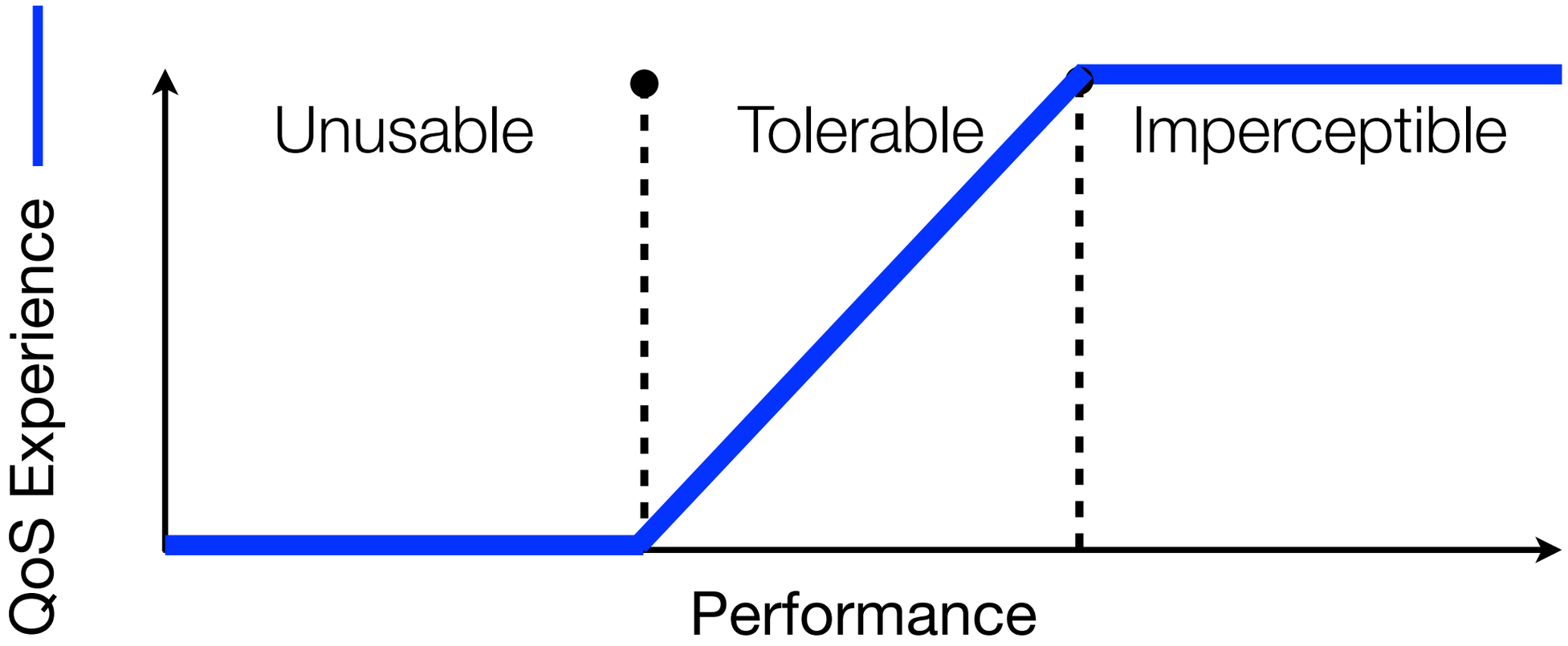
[OSDI 1996] Y. Endo et al., “Using Latency to Evaluate Interactive System Performance.”

Understanding Mobile Web QoS



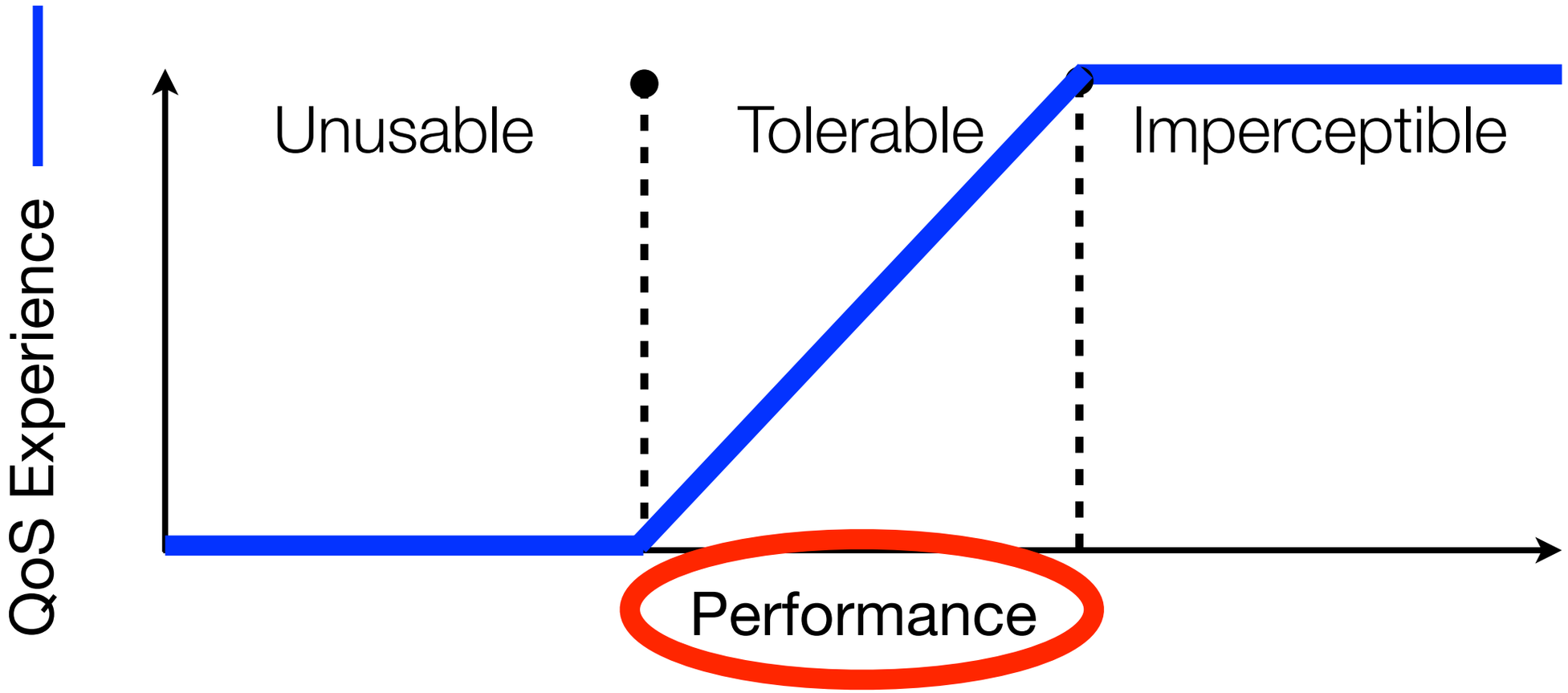
[OSDI 1996] Y. Endo et al., “Using Latency to Evaluate Interactive System Performance.”

Abstracting Mobile Web QoS



Abstracting Mobile Web QoS

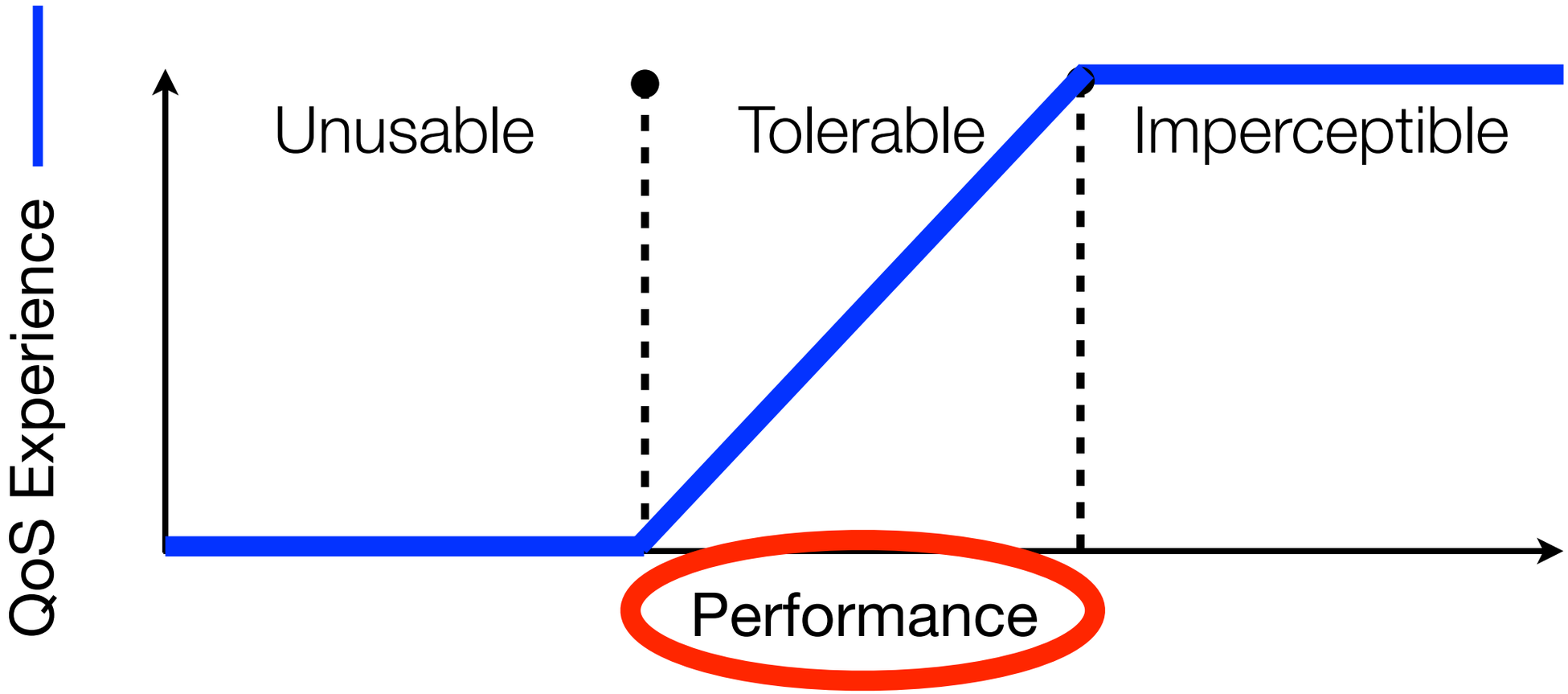
- **Performance metric**
  - Frame latency vs. Frame throughput



Abstracting Mobile Web QoS

- **Performance metric**
  - Frame latency vs. Frame throughput

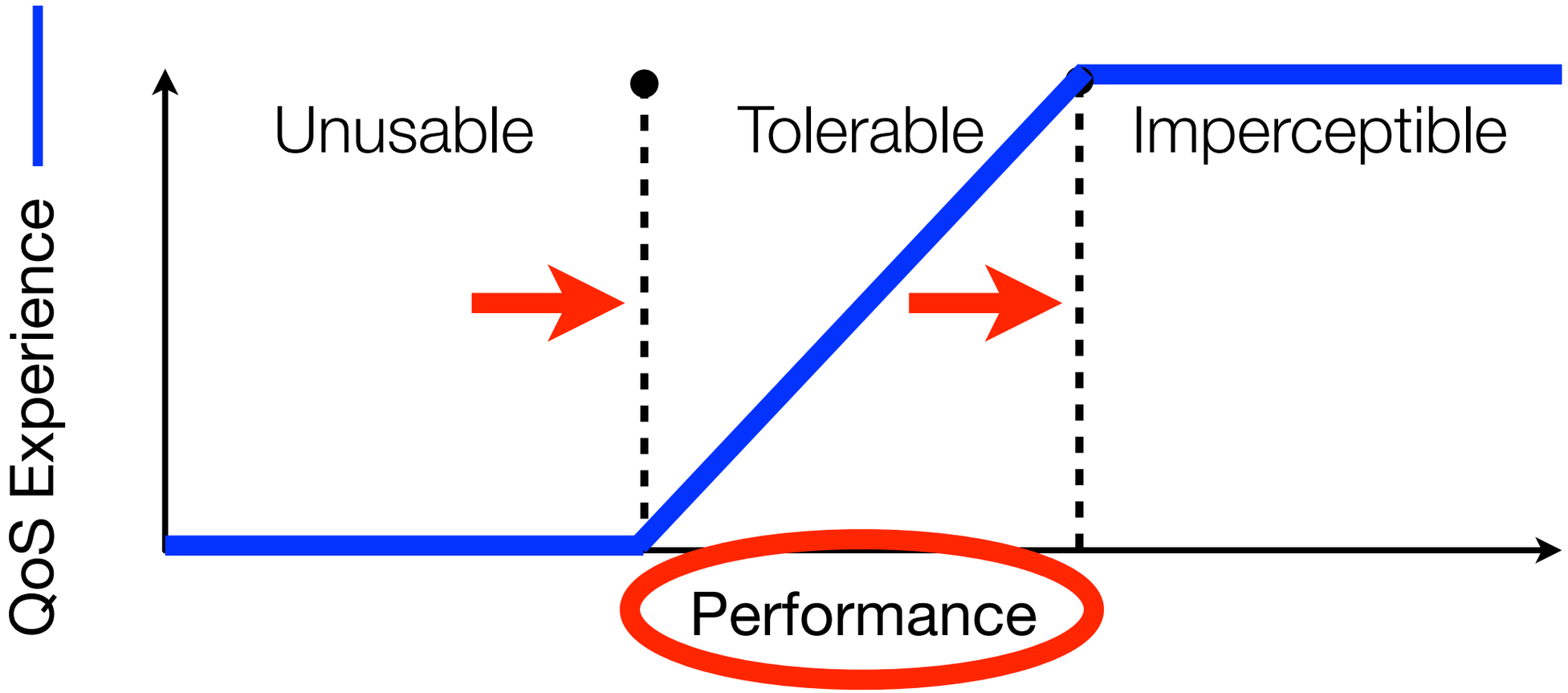
QoS Type



Abstracting Mobile Web QoS

- Performance metric
  - Frame latency vs. Frame throughput
- Threshold performance values
  - Imperceptible target vs. Usable target

QoS Type



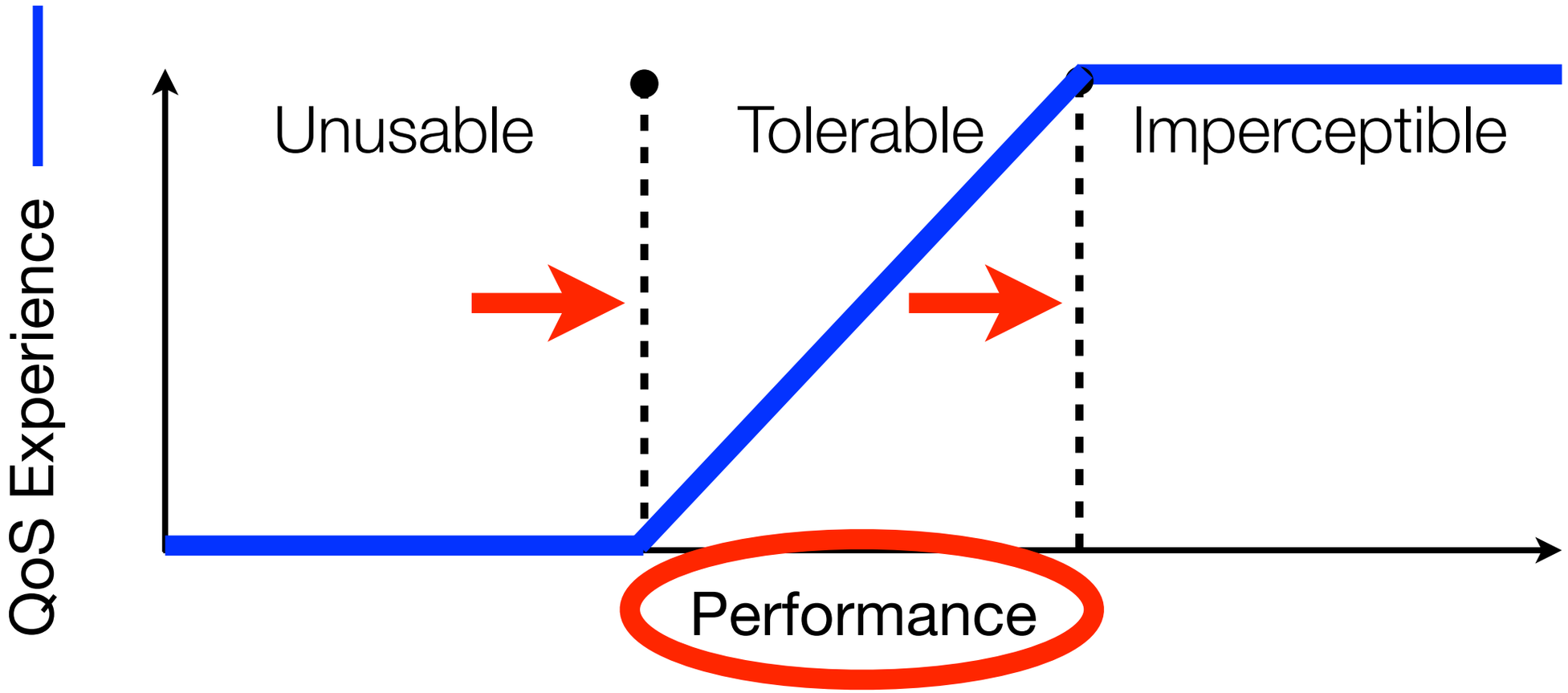
Abstracting Mobile Web QoS

- Performance metric
  - Frame latency vs. Frame throughput

QoS Type

- Threshold performance values
  - Imperceptible target vs. Usable target

QoS Target



Expressing Mobile Web QoS

Expressing Mobile Web QoS

```
<html> <head>
  <script>
    function animateMove() {
      /* Animation code omitted */
    }
  </script> </head> <body>
  <div ontouchend="animateMove()">
  <div/> <!– other elements -->
</body> </html>
```

Expressing Mobile Web QoS

element

```
<html> <head>
  <script>
    function animateMove() {
      /* Animation code omitted */
    }
  </script> </head> <body>
  <div ontouchend="animateMove()">
  <div/> <!– other elements -->
</body> </html>
```

Expressing Mobile Web QoS

element **event**

```
<html> <head>
  <script>
    function animateMove() {
      /* Animation code omitted */
    }
  </script> </head> <body>
  <div ontouchend=“animateMove()”>
  <div/> <!– other elements -->
</body> </html>
```

Expressing Mobile Web QoS

element **event**

```
<html> <head>
  <script>
    function animateMove() {
      /* Animation code omitted */
    }
  </script> </head> <body>
  <div ontouchend="animateMove()">
  <div/> <!– other elements -->
</body> </html>
```

Expressing Mobile Web QoS

element **event**

```
<html> <head>
  }
</script> </head> <body>
<div ontouchend="animateMove()">
<div/> <!– other elements -->
</body> </html>
```

Expressing QoS at an event granularity

Expressing Mobile Web QoS

Annotation

```
                element   event

<html> <head>
  <style>


  </style>
  <script>
   function animateMove() {
     /* Animation code omitted */
   }
  </script> </head> <body>
  <div ontouchend="animateMove()">
  <div/> <!— other elements -->
</body> </html>
```

Expressing Mobile Web QoS

Annotation

```
                element   event

<html> <head>
  <style>
    div {
      ontouchend
    }
  </style>
  <script>
   function animateMove() {
     /* Animation code omitted */
   }
  </script> </head> <body>
  <div ontouchend="animateMove()">
  <div/> <!— other elements -->
</body> </html>
```

Expressing Mobile Web QoS

Annotation

```
element {event: Type, Target}
```

```
<html> <head>
  <style>
    div {
      ontouchend: throughput, low;
    }
  </style>
  <script>
   function animateMove() {
     /* Animation code omitted */
   }
  </script> </head> <body>
  <div ontouchend="animateMove()">
  <div/> <!– other elements -->
</body> </html>
```

Expressing Mobile Web QoS


```
element {event: Type, Target}
```

Annotation

```
<html> <head>
  <style>
    div {
      ontouchend: throughput, low;
    }
  </style>
  <script>
    function animateMove() {
      /* Animation code omitted */
    }
  </script> </head> <body>
  <div ontouchend="animateMove()">
  <div/> <!– other elements -->
</body> </html>
```

Expressing Mobile Web QoS


```
element {event: Type, Target}
```

Annotation

```
<html> <head>
  <style>
    div {
      ontouchend: throughput, low;
    }
  </style>
  <script>
    function animateMove() {
      /* Animation code omitted */
    }
  </script> </head> <body>
  <div ontouchend="animateMove()">
  <div/> <!– other elements -->
</body> </html>
```

Expressing Mobile Web QoS


```
element {event: Type, Target}
```

Annotation

```
<html> <head>
  <style>
    div {
      ontouchend: throughput, low;
    }
  </style>
  <script>
    function newAnimateMove() {
      /* New animation code */
    }
  </script> </head> <body>
  <div ontouchend="animateMove()">
  <div/> <!– other elements -->
</body> </html>
```

Expressing Mobile Web QoS


```
element {event: Type, Target}
```

Annotation

```
<html> <head>
  <style>
    div {
      ontouchend: throughput, low;
    }
  </style>
  <script>
    function newAnimateMove() {
      /* New animation code */
    }
  </script> </head> <body>
  <div ontouchend="animateMove()">
  <div/> <!– other elements -->
</body> </html>
```

Implementation independent

Expressing Mobile Web QoS


```
element {event: Type, Target}
```

Annotation

```
<html> <head>
  <style>
    div {
      ontouchend: throughput, low;
    }
  </style>
  <script>
    function newAnimateMove() {
      /* New animation code */
    }
  </script> </head> <body>
  <div ontouchend="animateMove()">
  <div/> <!– other elements -->
</body> </html>
```

Implementation independent

Expressing Mobile Web QoS

CSS
Cascading Style Sheet

```
element {event: Type, Target}
```

Annotation

```
<html> <head>
  <style>
    div {
      ontouchend: throughput, low;
    }
  </style>
  <script>
    function newAnimateMove() {
      /* New animation code */
    }
  </script> </head> <body>
  <div ontouchend="animateMove()">
  <div/> <!– other elements -->
</body> </html>
```

Implementation independent

Non-interfering w.r.t. functionality

GreenWeb Annotation Process

Original application

Manual Annotation

GreenWeb-annotated application

GreenWeb Annotation Process

Original application

Automatic Annotation?

GreenWeb-annotated application

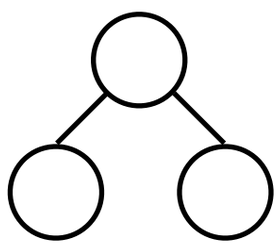
GreenWeb Annotation Process

- **AutoGreen**: *automatically* reasons about and inserts **GreenWeb** annotations

GreenWeb Annotation Process

- **AutoGreen**: *automatically* reasons about and inserts **GreenWeb** annotations

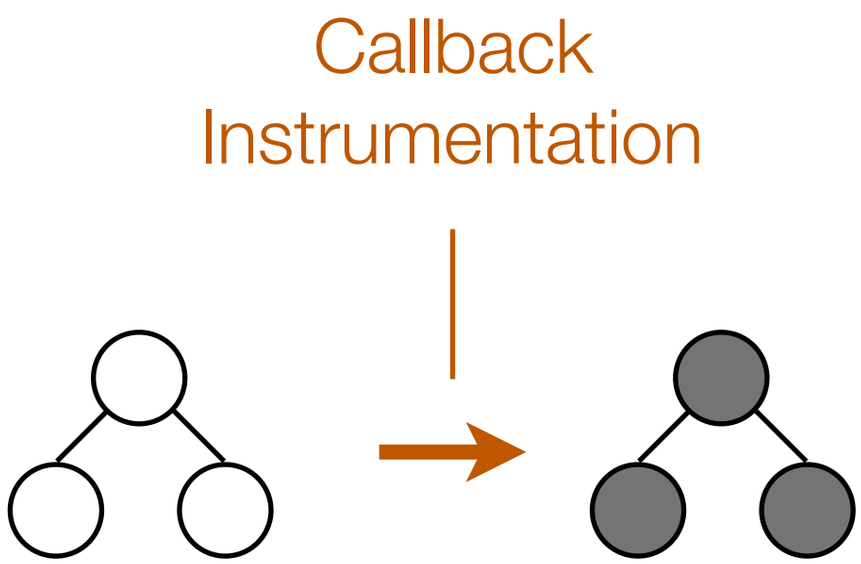
GreenWeb Annotation Process



GreenWeb-annotated application

- **AutoGreen**: *automatically* reasons about and inserts **GreenWeb** annotations

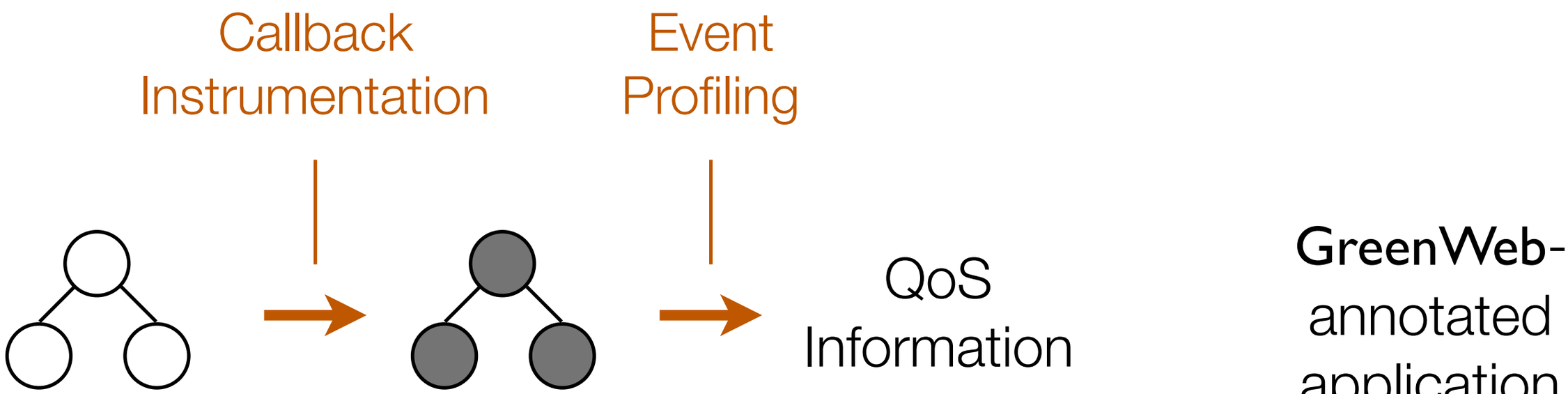
GreenWeb Annotation Process



GreenWeb-annotated application

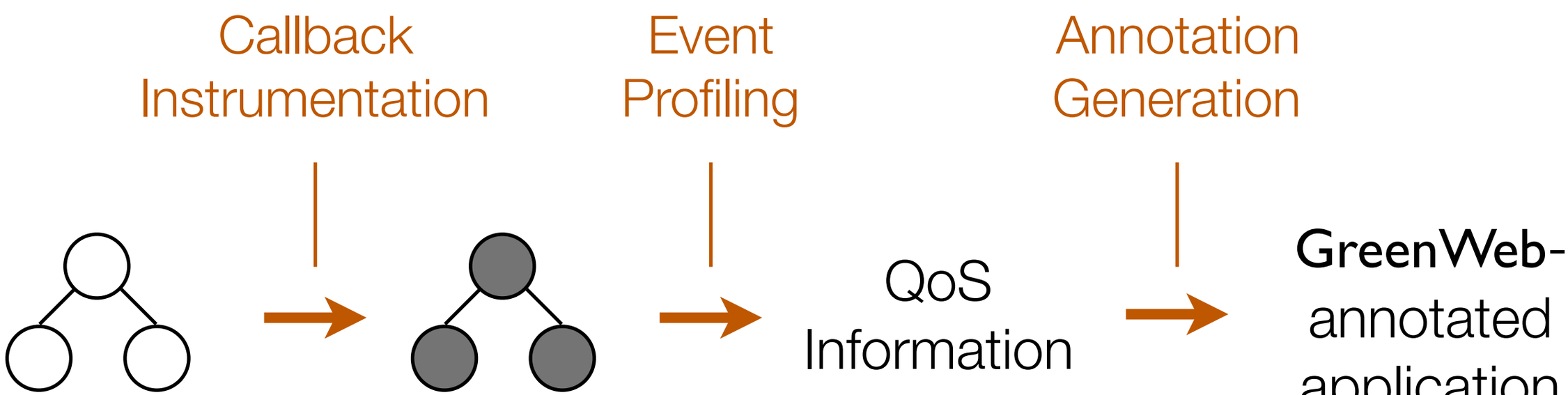
- **AutoGreen**: *automatically* reasons about and inserts **GreenWeb** annotations

GreenWeb Annotation Process



- **AutoGreen**: *automatically* reasons about and inserts **GreenWeb** annotations

GreenWeb Annotation Process



- **AutoGreen**: *automatically* reasons about and inserts **GreenWeb** annotations

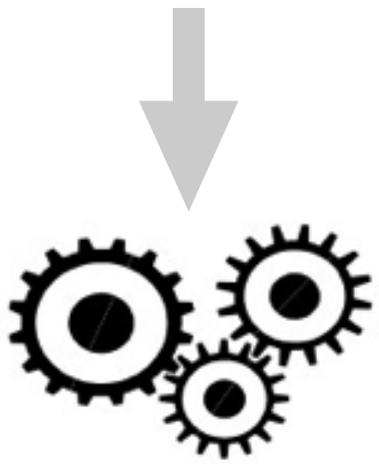
GreenWeb: Language for Energy-Efficiency

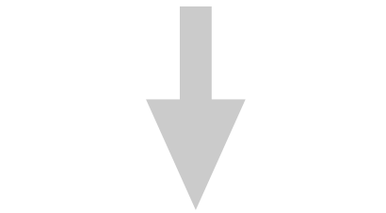
- **Language abstractions** for expressing QoS

- Runtime
  the QoS constraints

- Result
  hardware/software implementations

GreenWeb: Language for Energy-Efficiency





- Language abstractions
- **Runtime** that saves energy while meeting the QoS constraints
- Result
  hardware/software implementations

GreenWeb Runtime Overview

GreenWeb Runtime Overview

GreenWeb Runtime Overview

QoS
Annotations

GreenWeb Runtime Overview

Runtime Objective

Enforcing event-level QoS at the frame-level energy-efficiently

QoS Annotations

GreenWeb Runtime Overview

Runtime Objective

Enforcing event-level QoS at the frame-level energy-efficiently

QoS type: latency

QoS target: 16 ms

GreenWeb Runtime Overview

Runtime Objective

Enforcing event-level QoS at the frame-level energy-efficiently

QoS type: latency

QoS target: 16 ms

GreenWeb Runtime Overview

Runtime Objective

Enforcing event-level QoS at the frame-level energy-efficiently

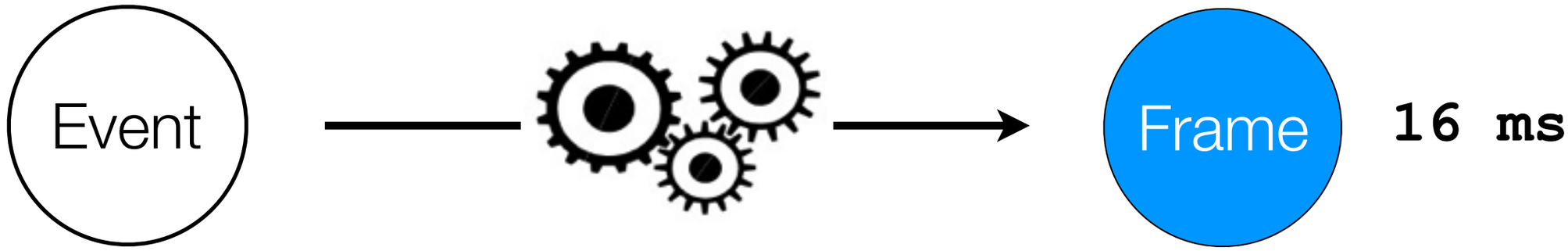
QoS type: latency

QoS target: 16 ms

GreenWeb Runtime Overview

Runtime Objective

Enforcing event-level QoS at the frame-level energy-efficiently



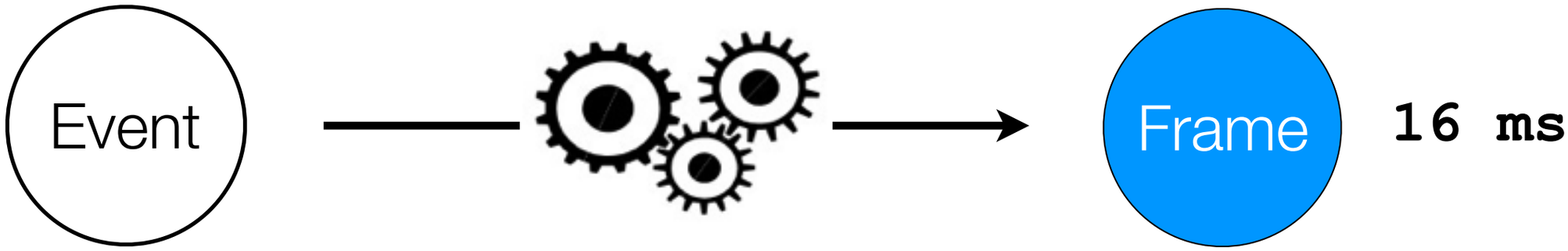
QoS type: latency

QoS target: 16 ms

GreenWeb Runtime Overview

Runtime Objective

Enforcing event-level QoS at the frame-level energy-efficiently



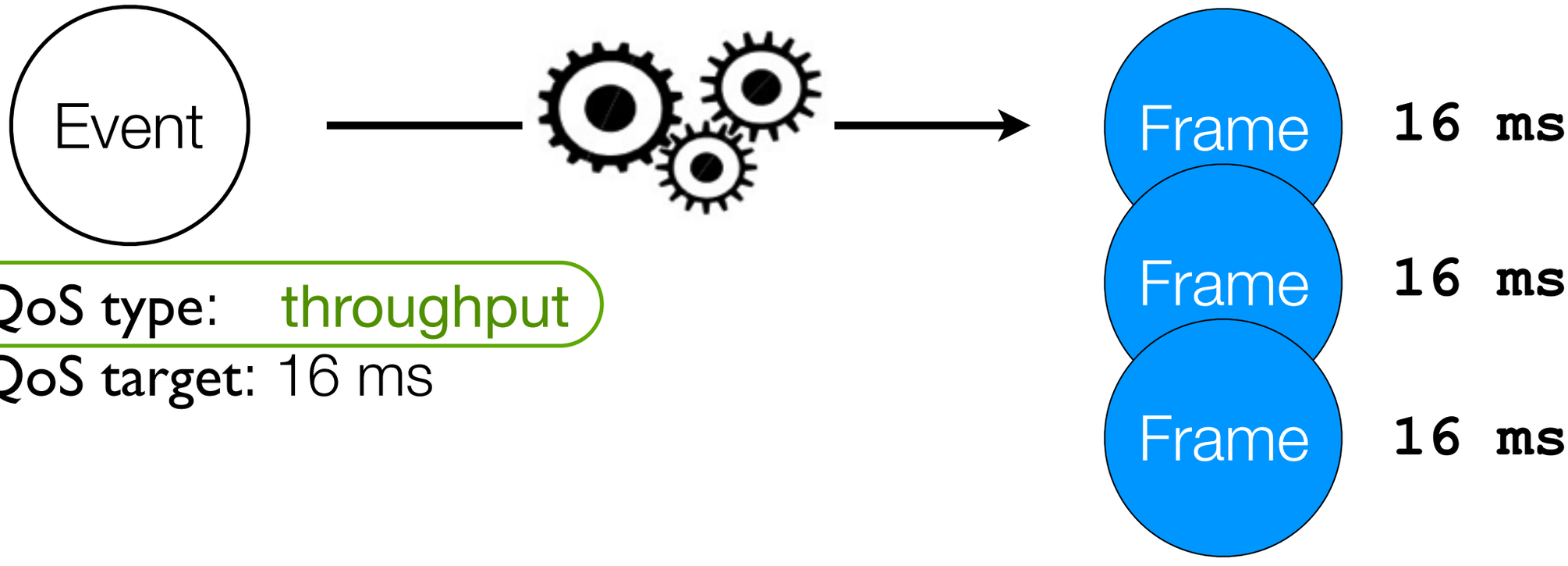
QoS type: throughput

QoS target: 16 ms

GreenWeb Runtime Overview

Runtime Objective

Enforcing event-level QoS at the frame-level energy-efficiently



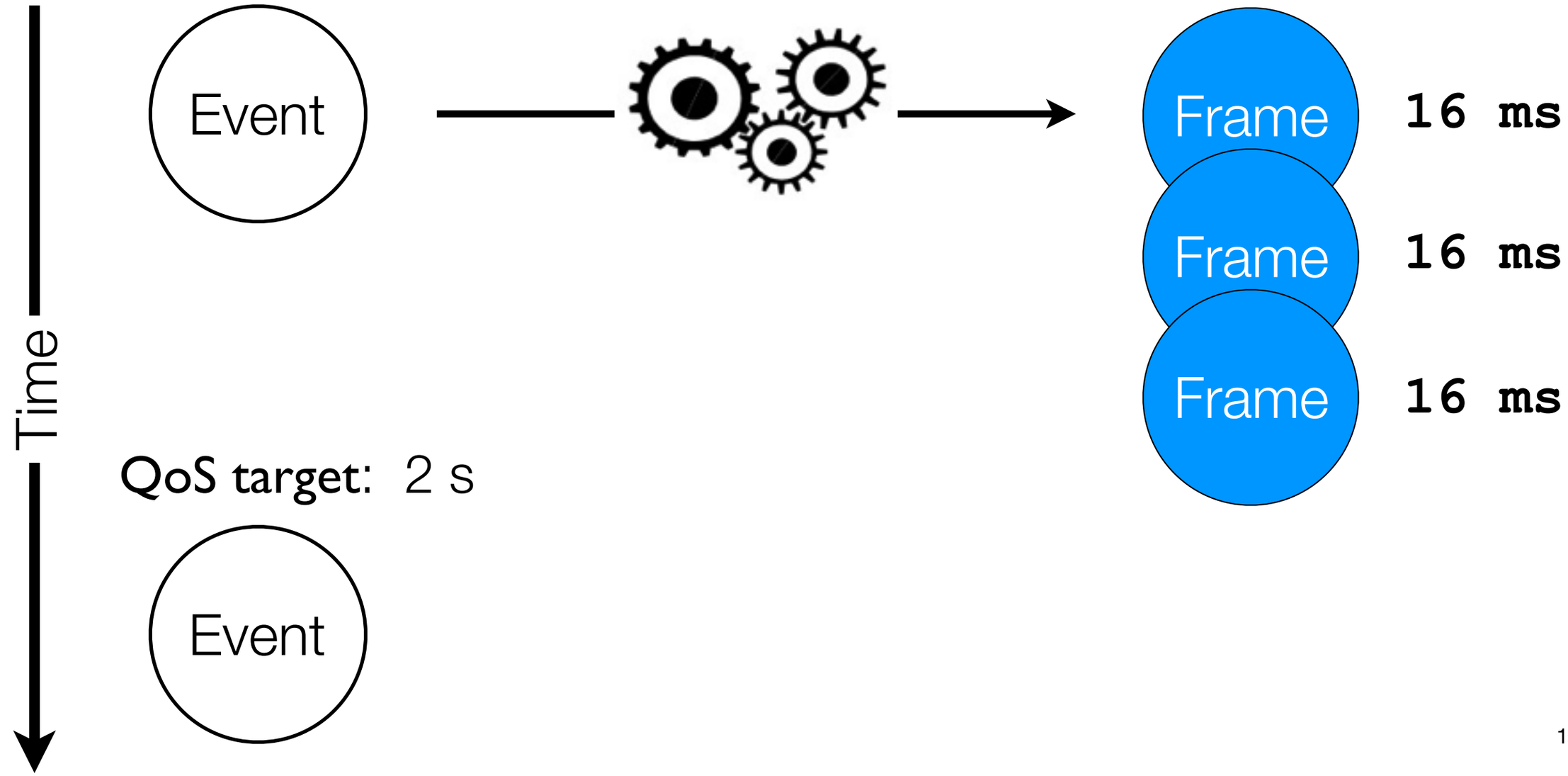
QoS type: throughput

QoS target: 16 ms

GreenWeb Runtime Overview

Runtime Objective

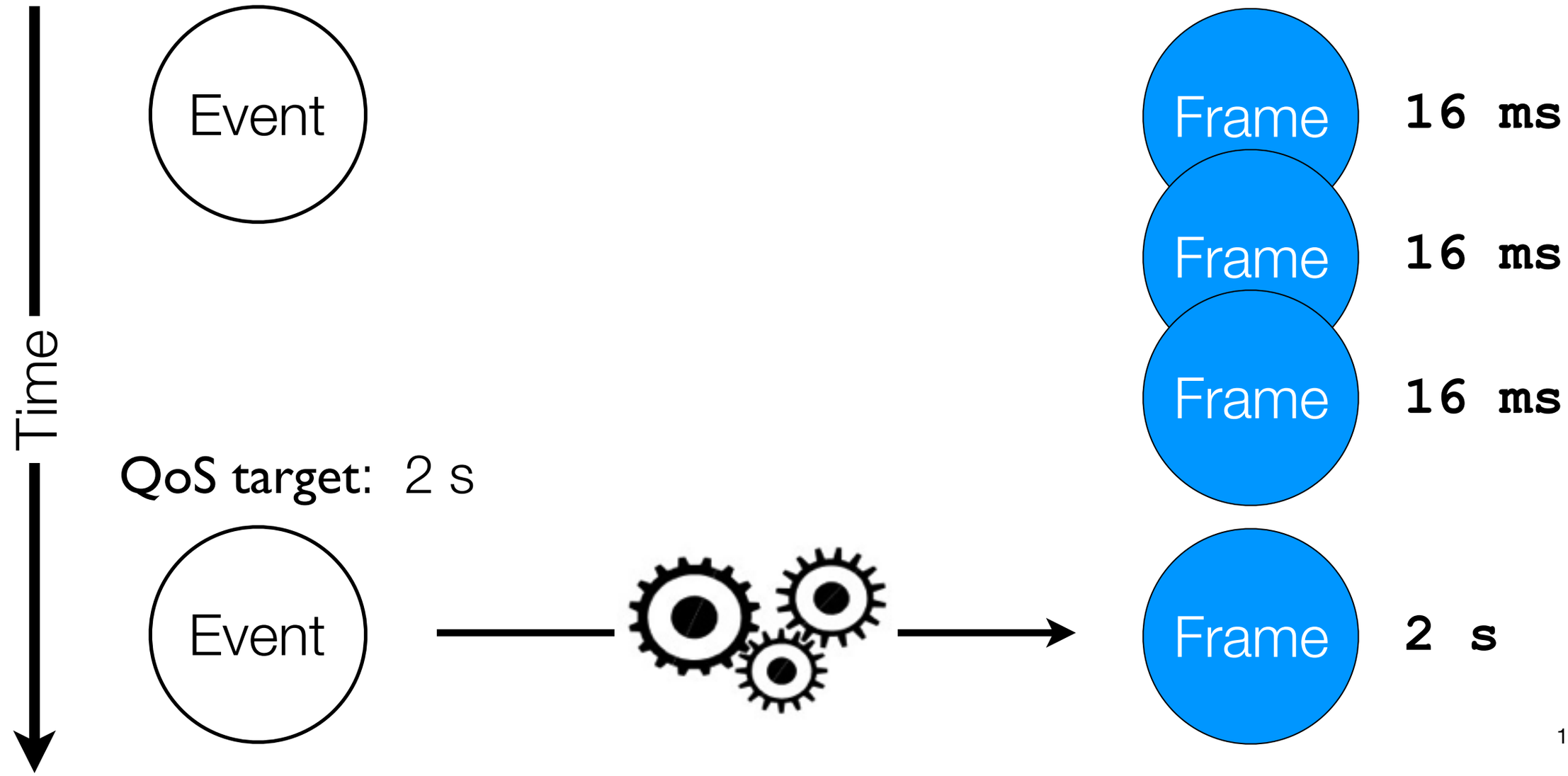
Enforcing event-level QoS at the frame-level energy-efficiently



GreenWeb Runtime Overview

Runtime Objective

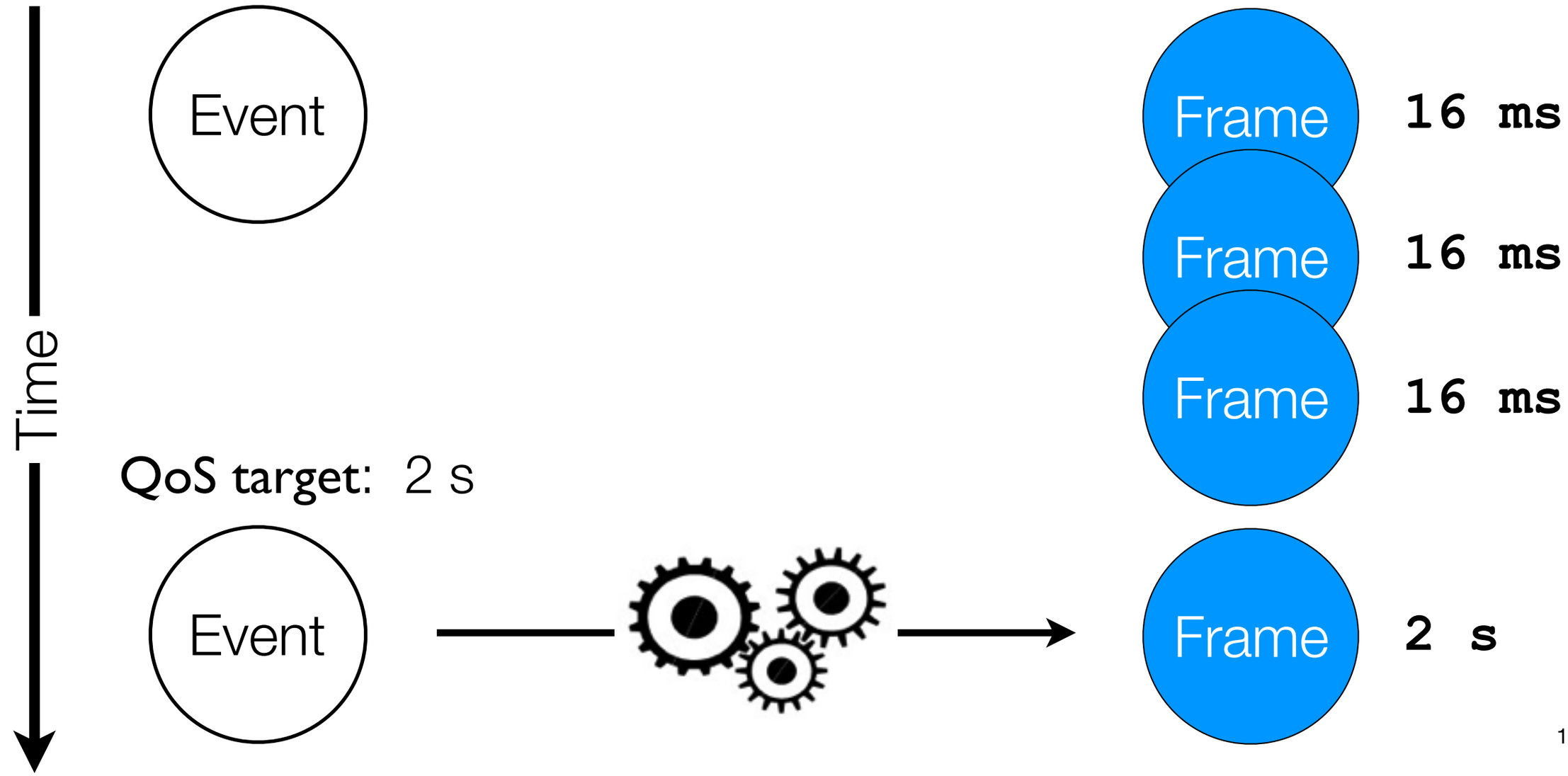
Enforcing event-level QoS at the frame-level energy-efficiently



GreenWeb Runtime Overview

Runtime Objective

Enforcing event-level QoS at the frame-level energy-efficiently {



GreenWeb Runtime Overview

Runtime Objective

Enforcing event-level QoS at the frame-level energy-efficiently

{

1. Frame Association

Time

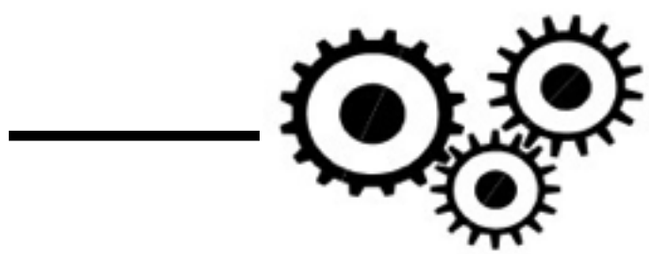
Event

Frame — 16 ms

Frame — 16 ms

Frame — 16 ms

Event



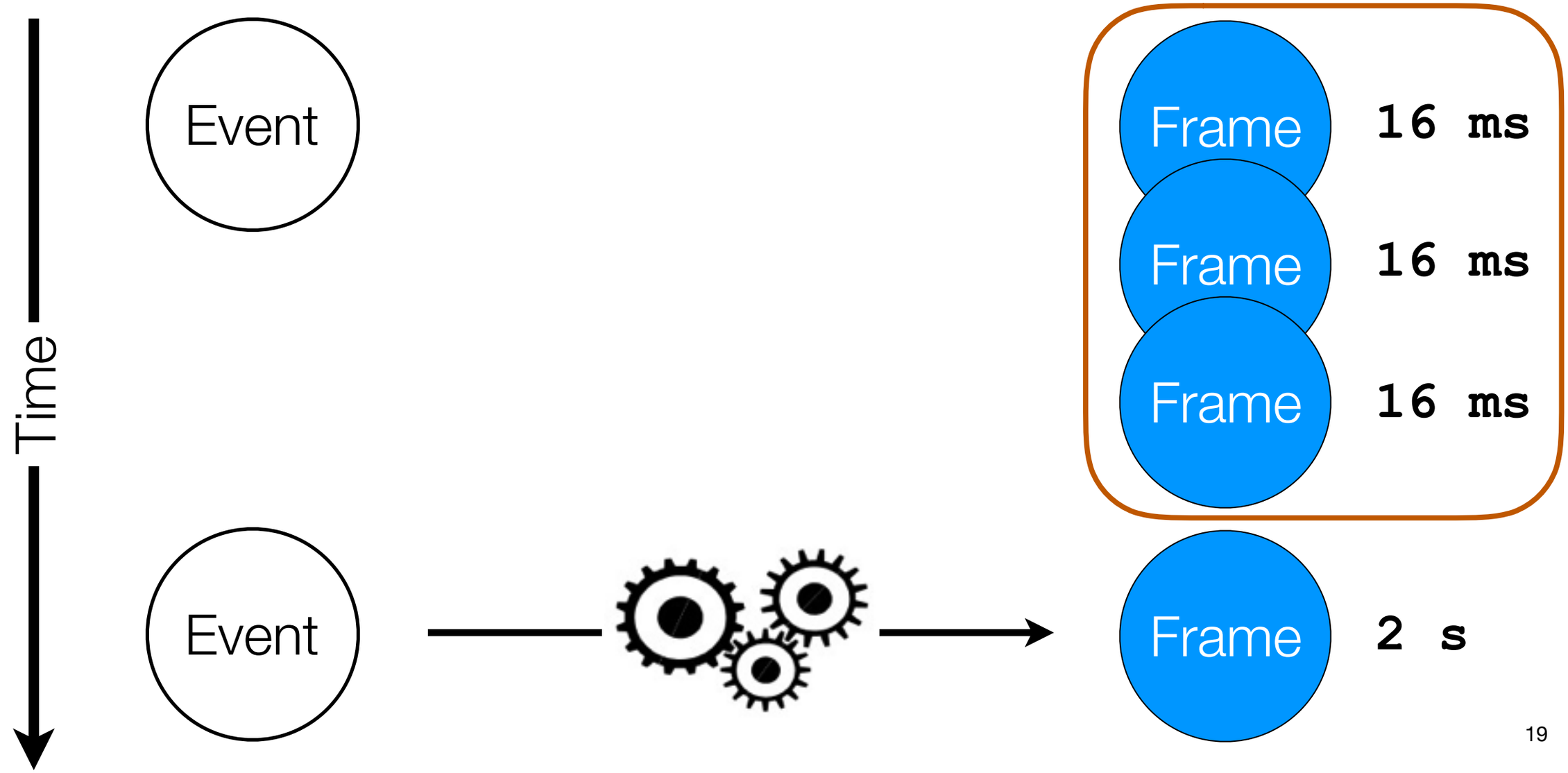
Frame — 2 s

GreenWeb Runtime Overview

Runtime Objective

Enforcing event-level QoS at the frame-level energy-efficiently

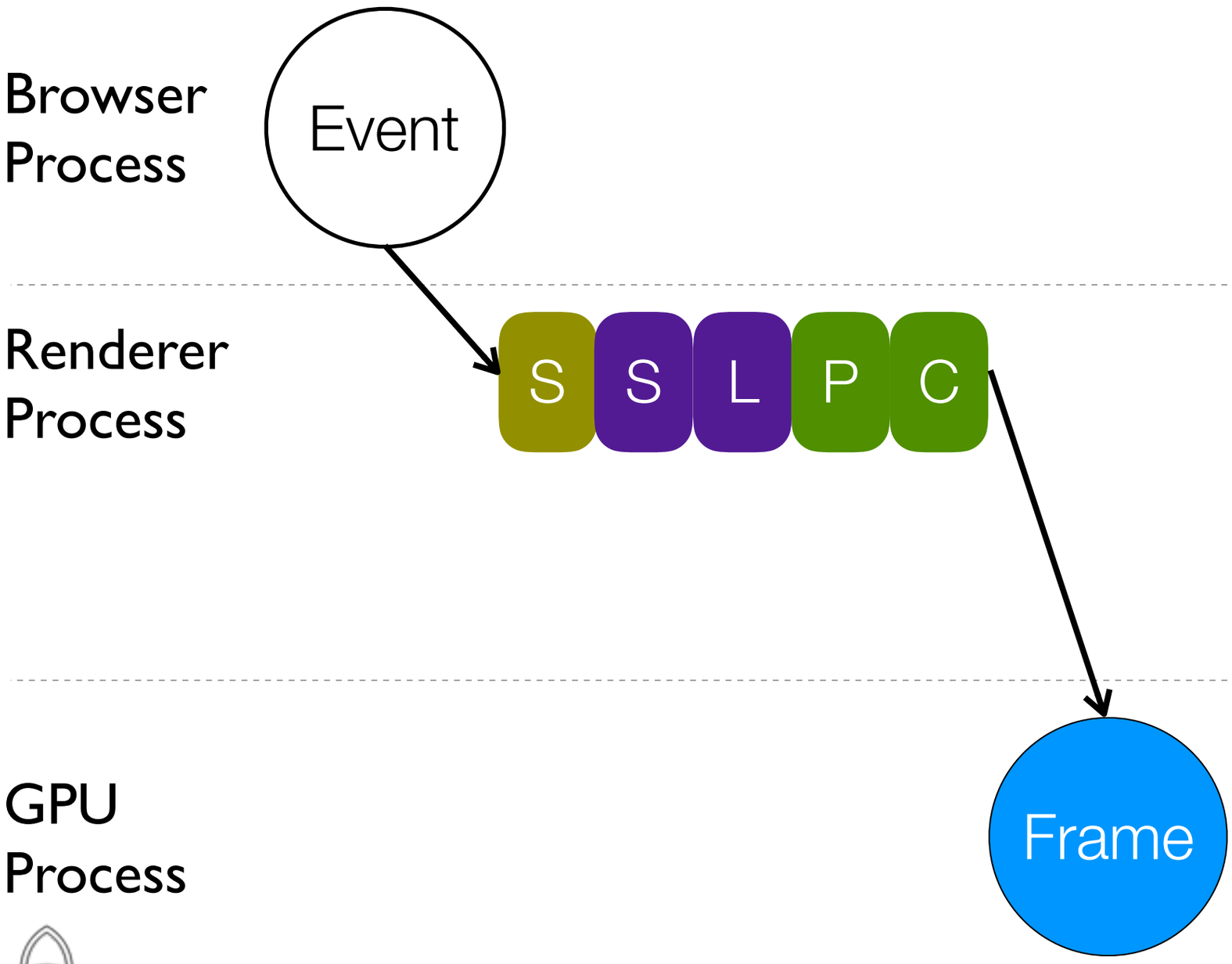
1. Frame Association
2. Frame Scheduling



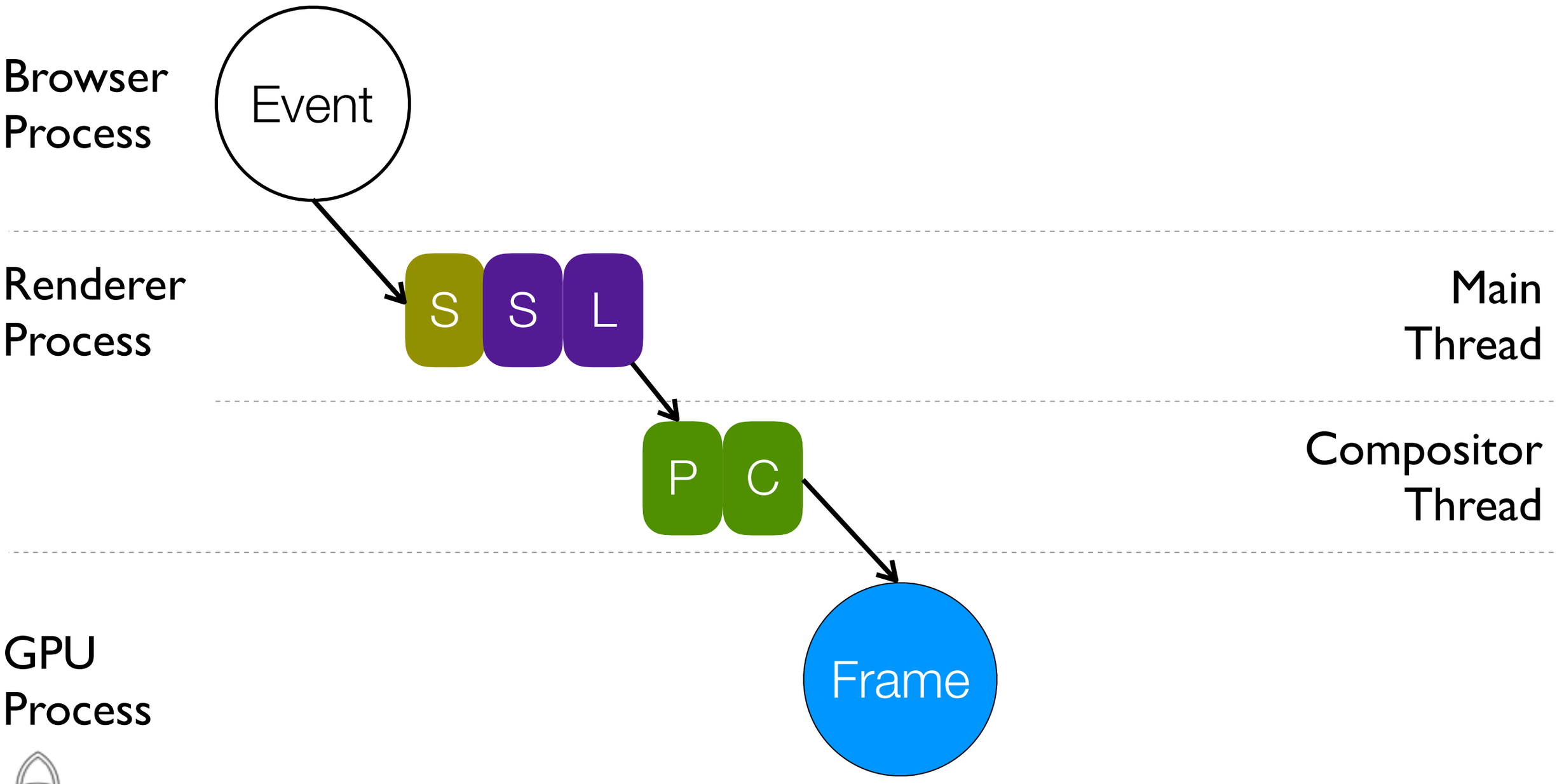
Frame Association

Frame Association

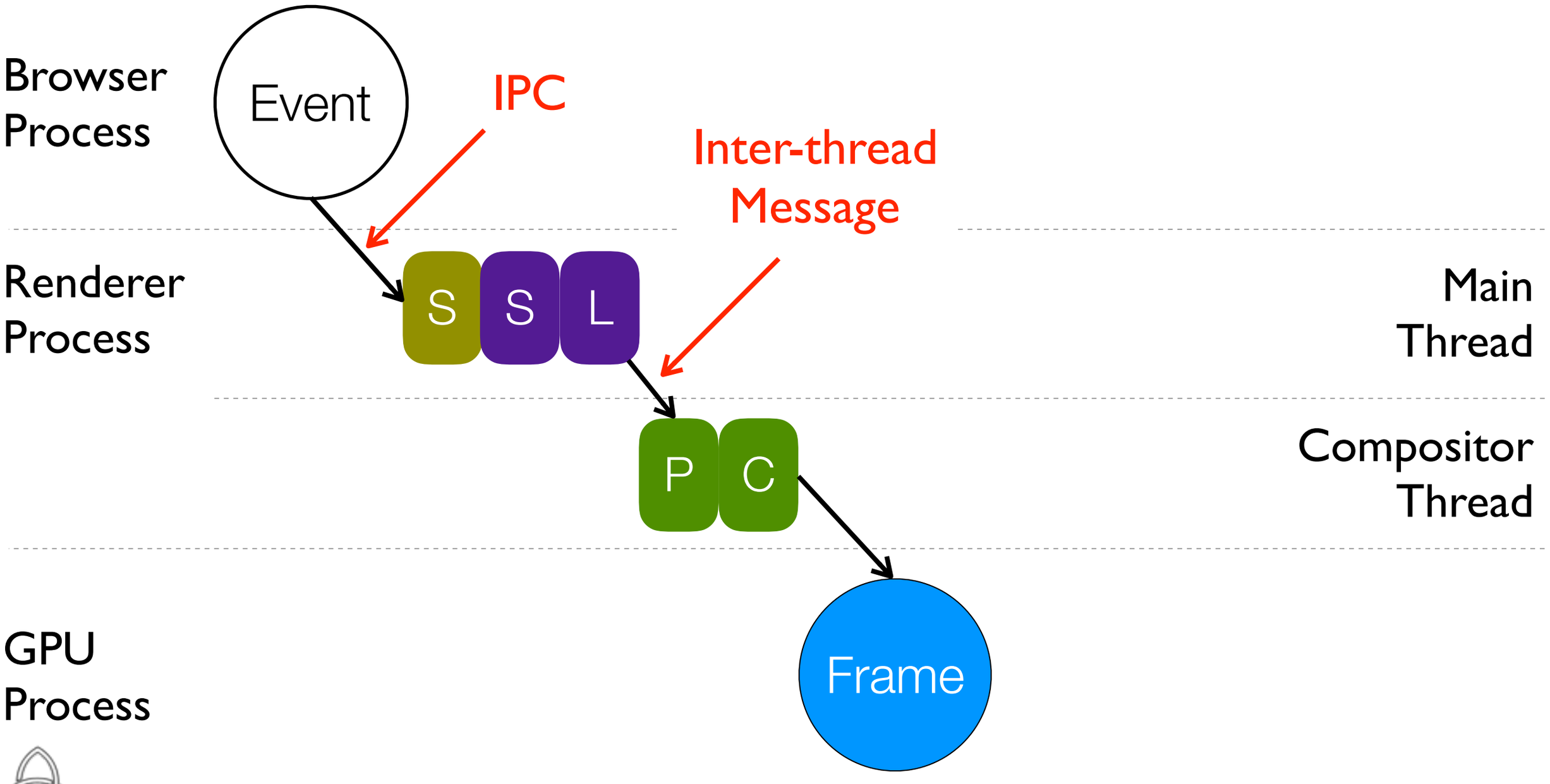
Frame Association



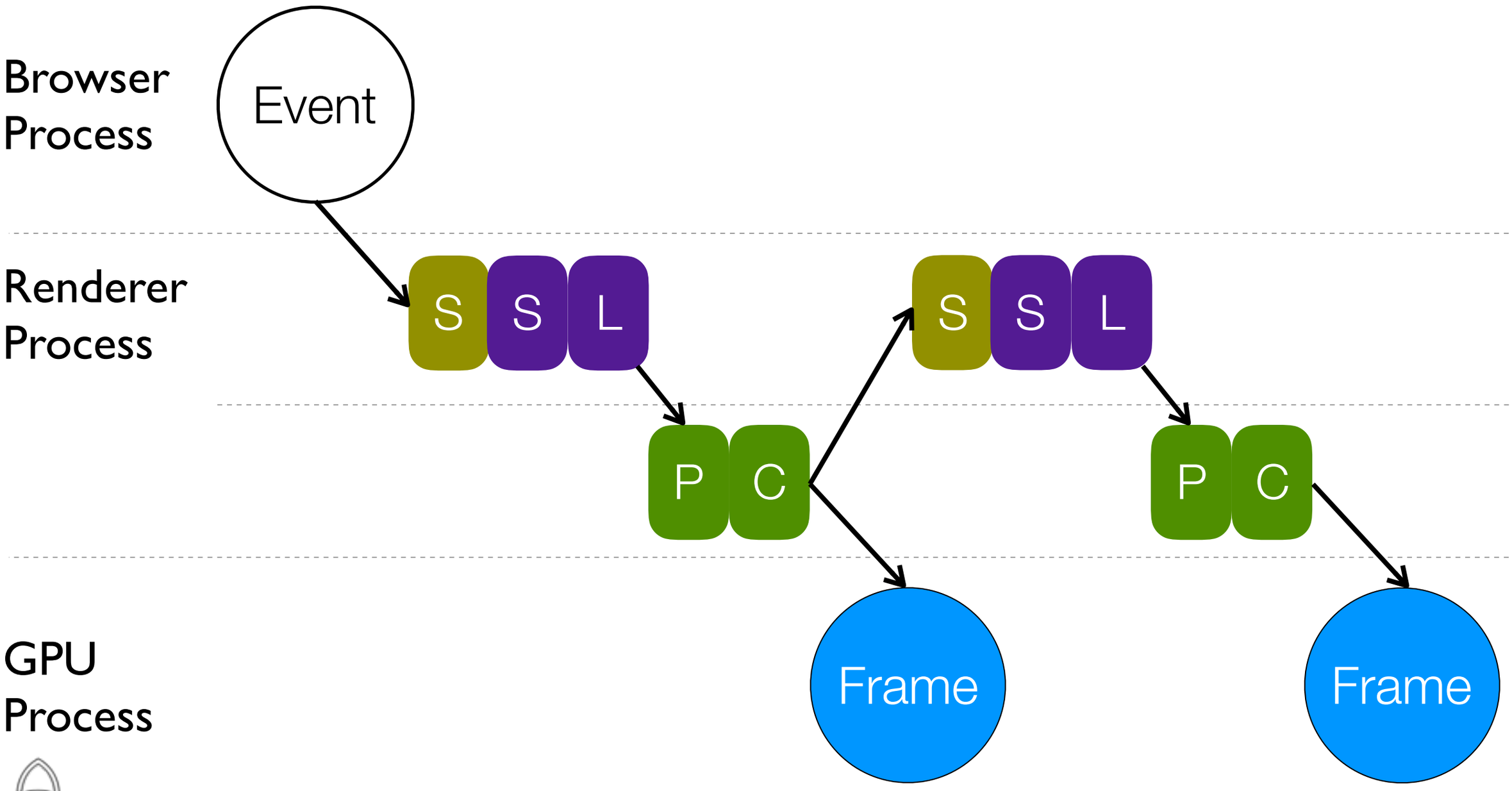
Frame Association



Frame Association

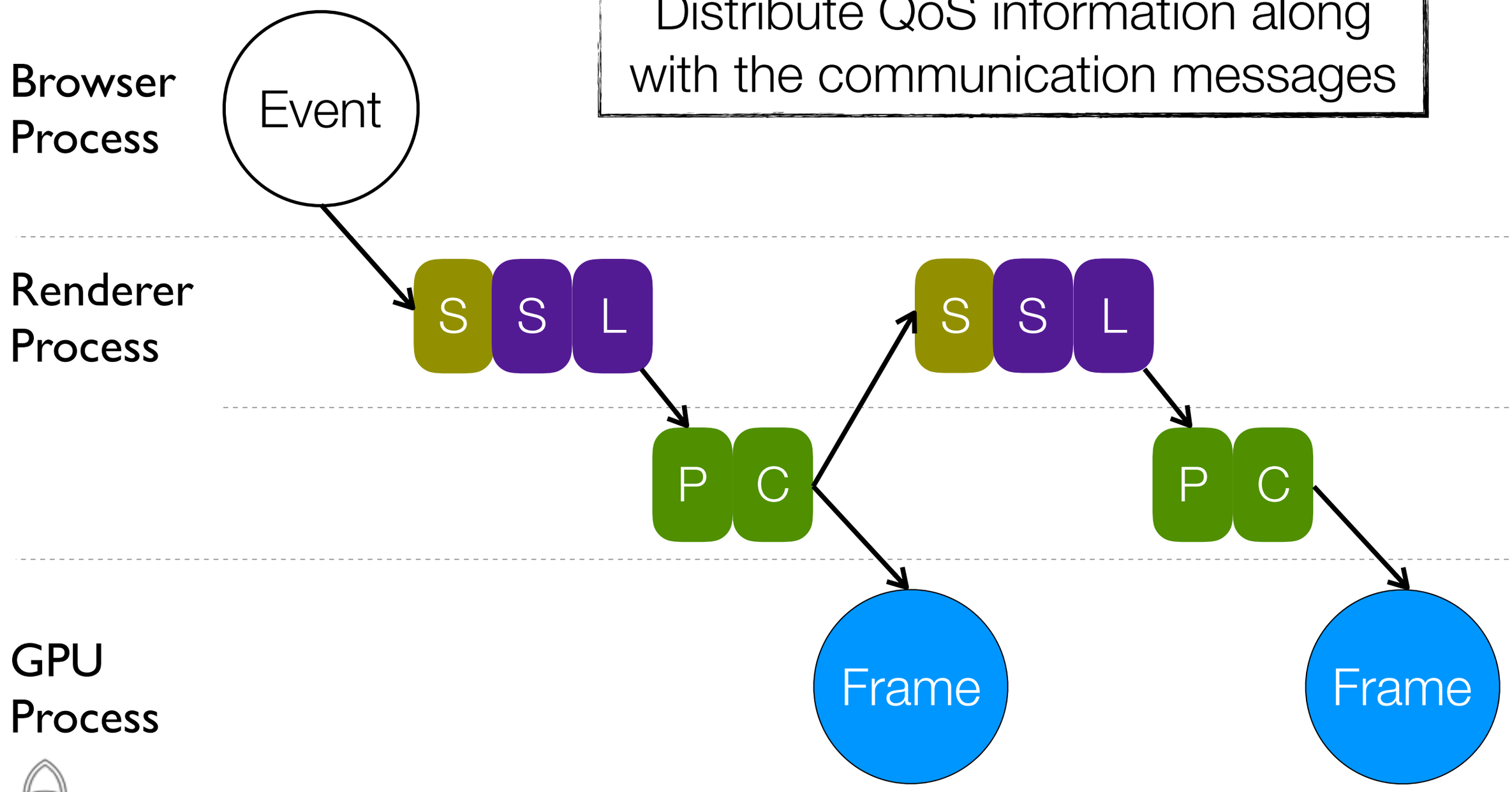


Frame Association



Frame Association

Distribute QoS information along with the communication messages



Choices of Energy-saving Techniques

GreenWeb can support a range of energy saving techniques

Choices of Energy-saving Techniques

GreenWeb can support a range of energy saving techniques

- Dynamic resolution scaling [MobiCom 2015]
- Power-saving display colors [MobiSys 2012]
- Selective resource loading [NSDI 2015]

Choices of Energy-saving Techniques

GreenWeb can support a range of energy saving techniques

- Dynamic resolution scaling [MobiCom 2015]
- Power-saving display colors [MobiSys 2012]
- Selective resource loading [NSDI 2015]
- ACMP-based hardware mechanism

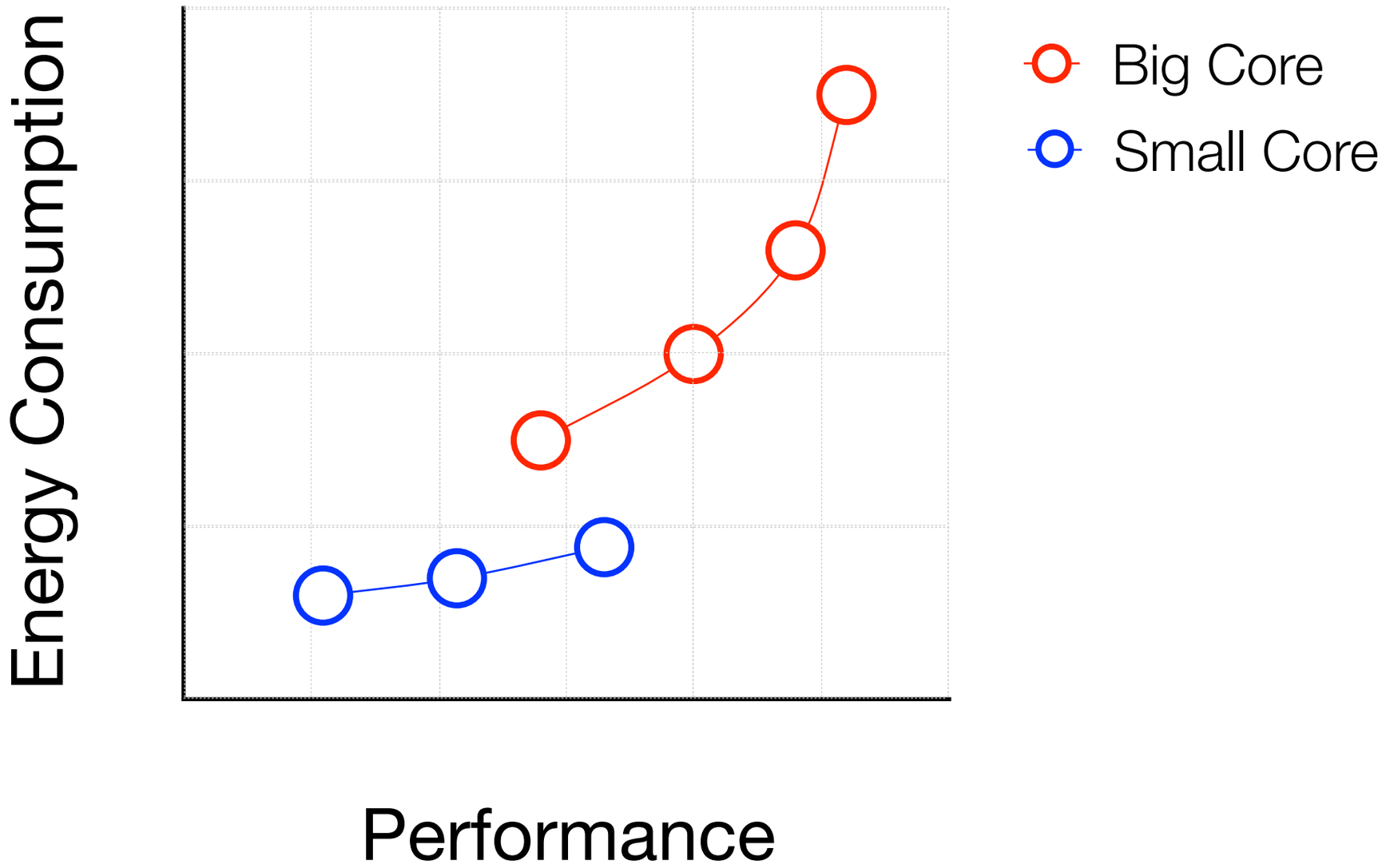
ACMP-based Hardware Substrate

ACMP-based Hardware Substrate

- Asymmetric Chip-multiprocessor, a.k.a., *Big/Little* architecture

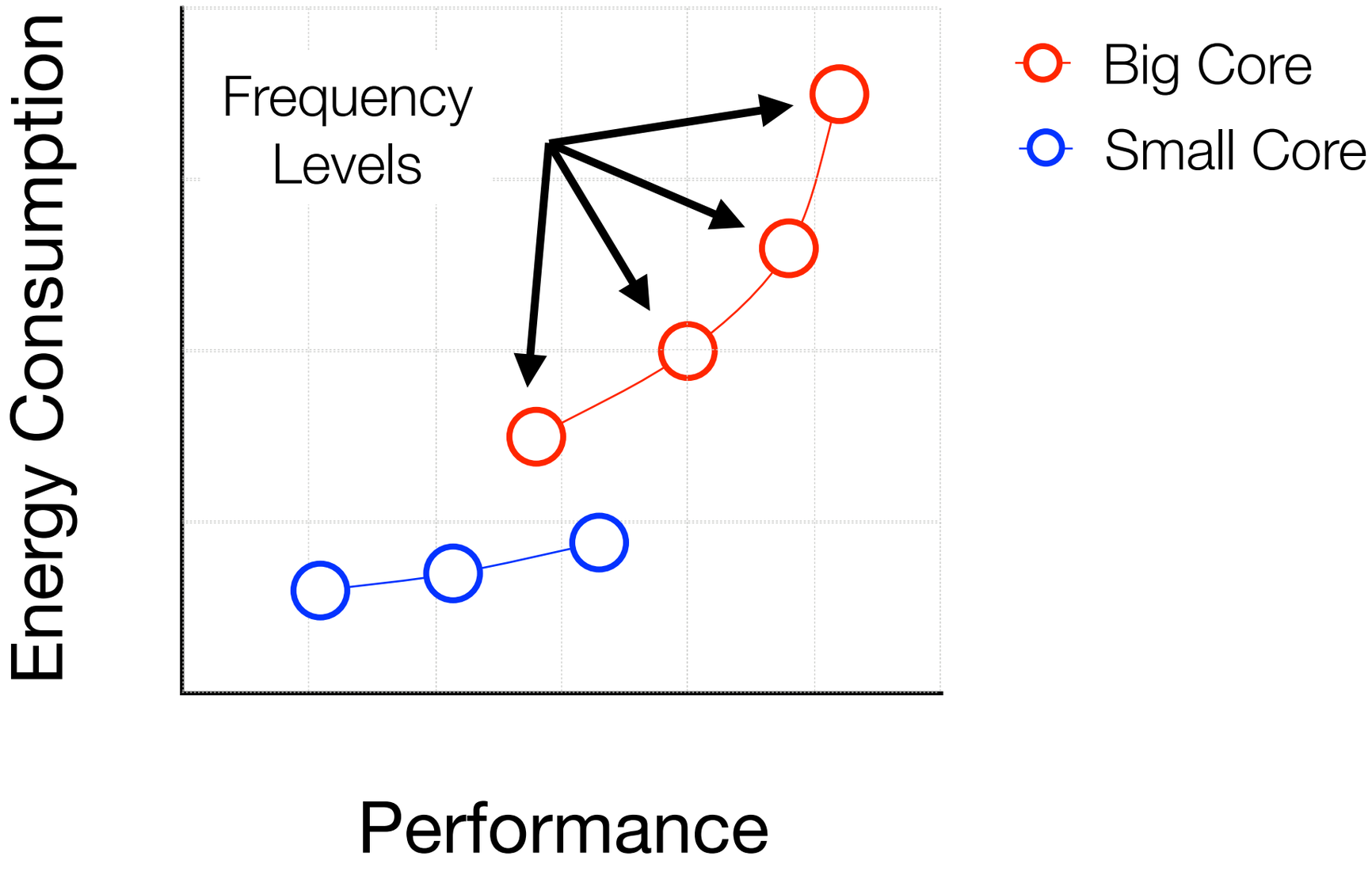
ACMP-based Hardware Substrate

- Asymmetric Chip-multiprocessor, a.k.a., *Big/Little* architecture



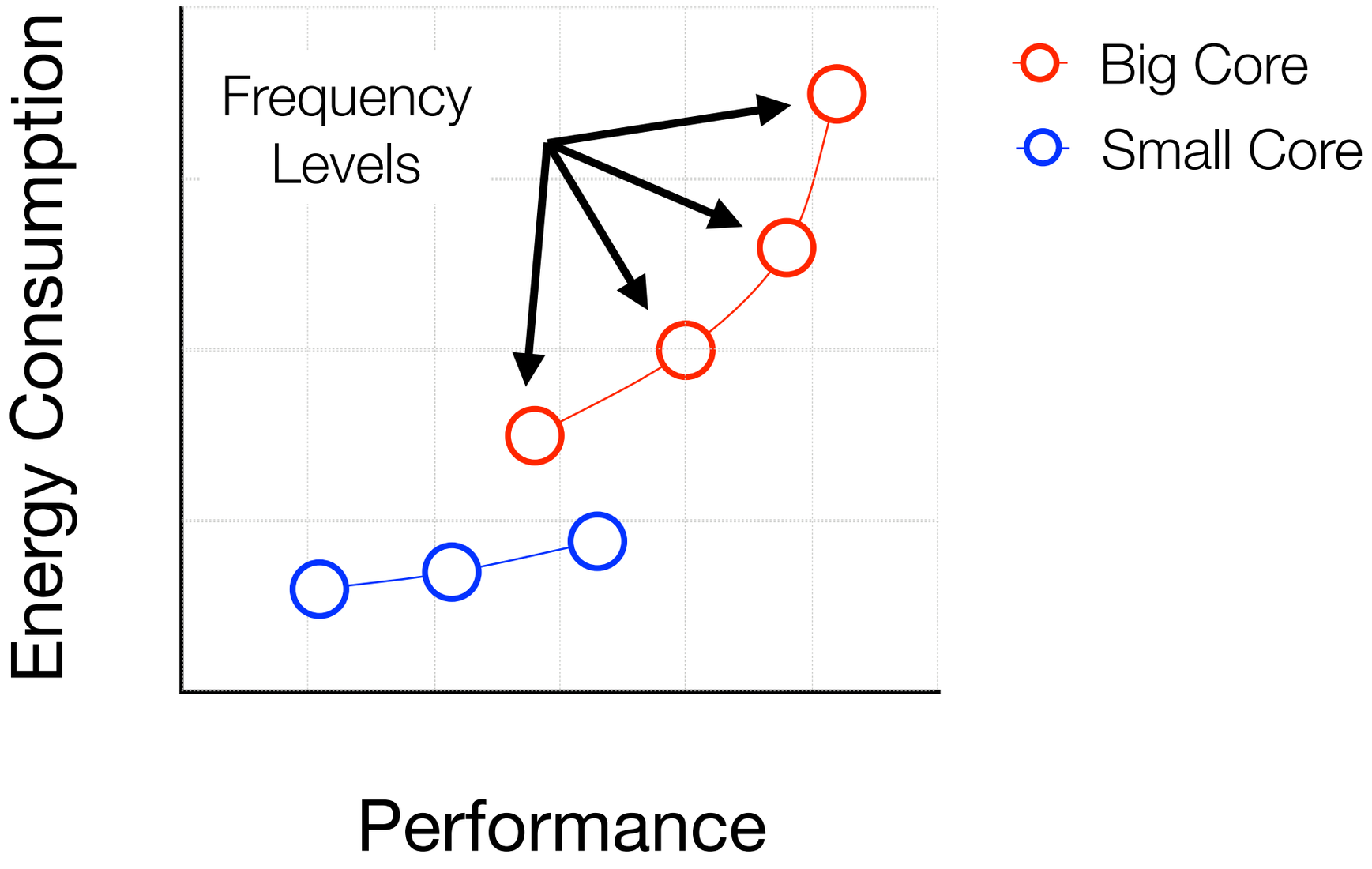
ACMP-based Hardware Substrate

- Asymmetric Chip-multiprocessor, a.k.a., *Big/Little* architecture

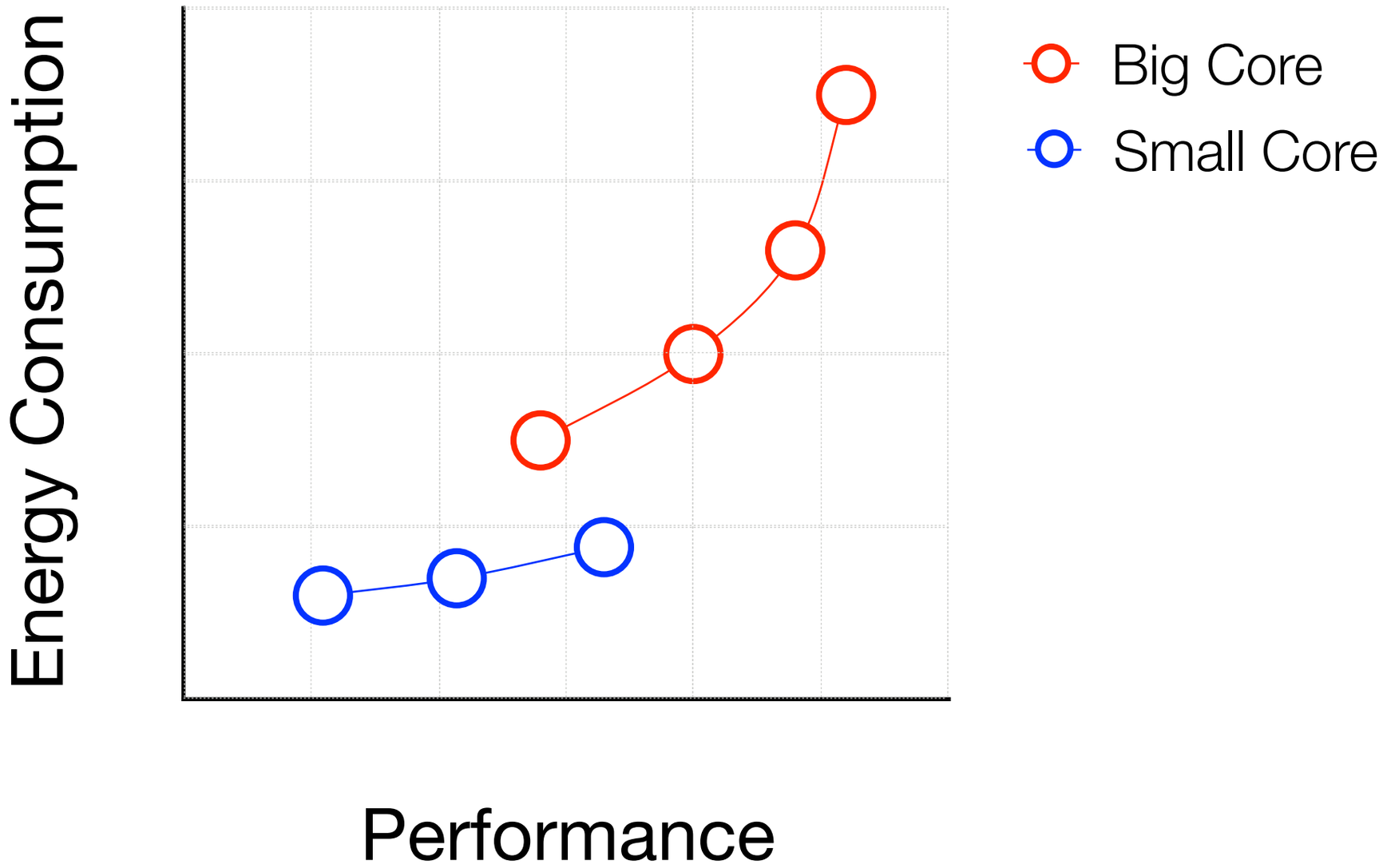


ACMP-based Hardware Substrate

- Asymmetric Chip-multiprocessor, a.k.a., *Big/Little* architecture
- Already used in commodity devices (e.g., Samsung Galaxy S6)

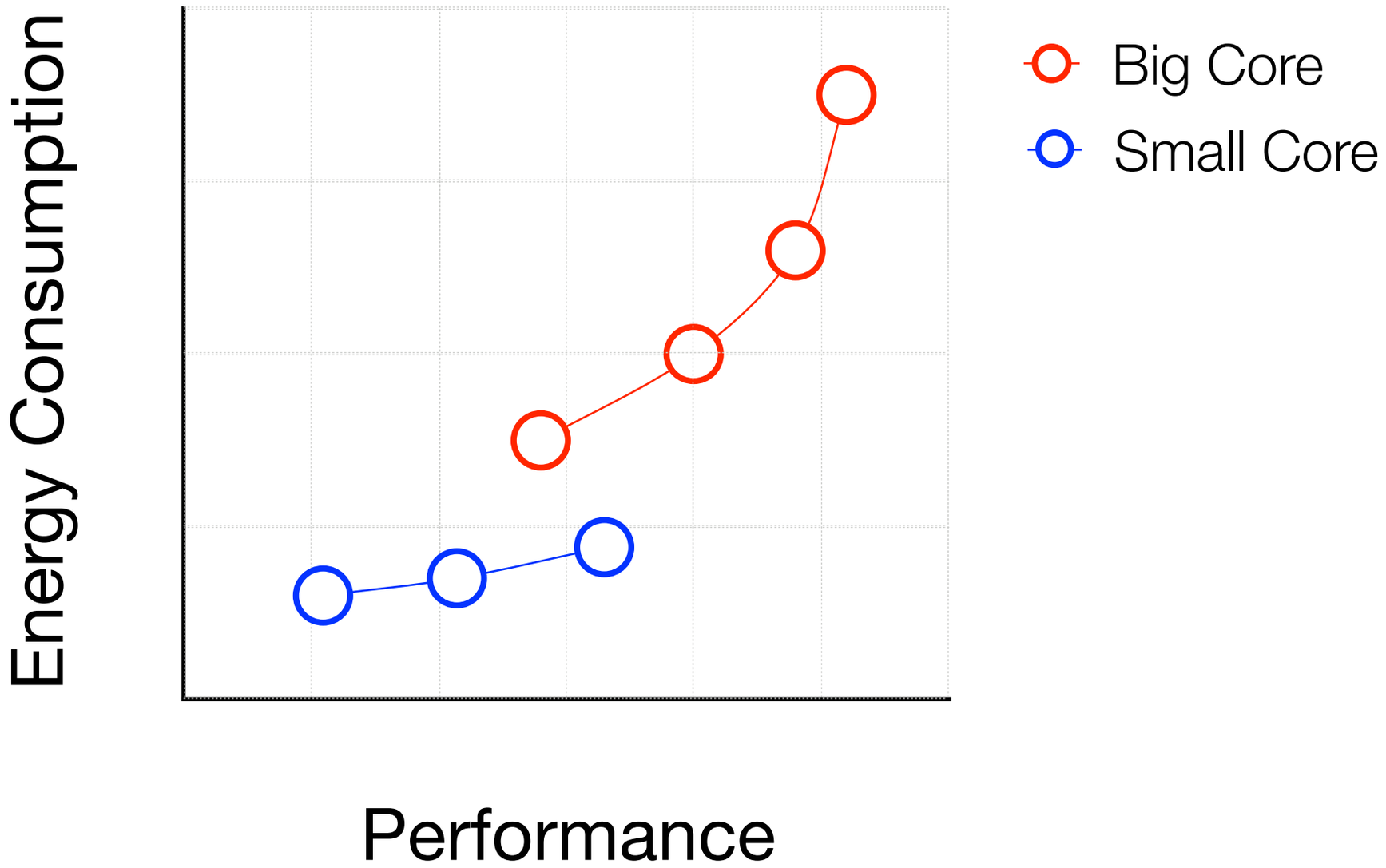


ACMP-based **GreenWeb** Runtime



ACMP-based **GreenWeb** Runtime

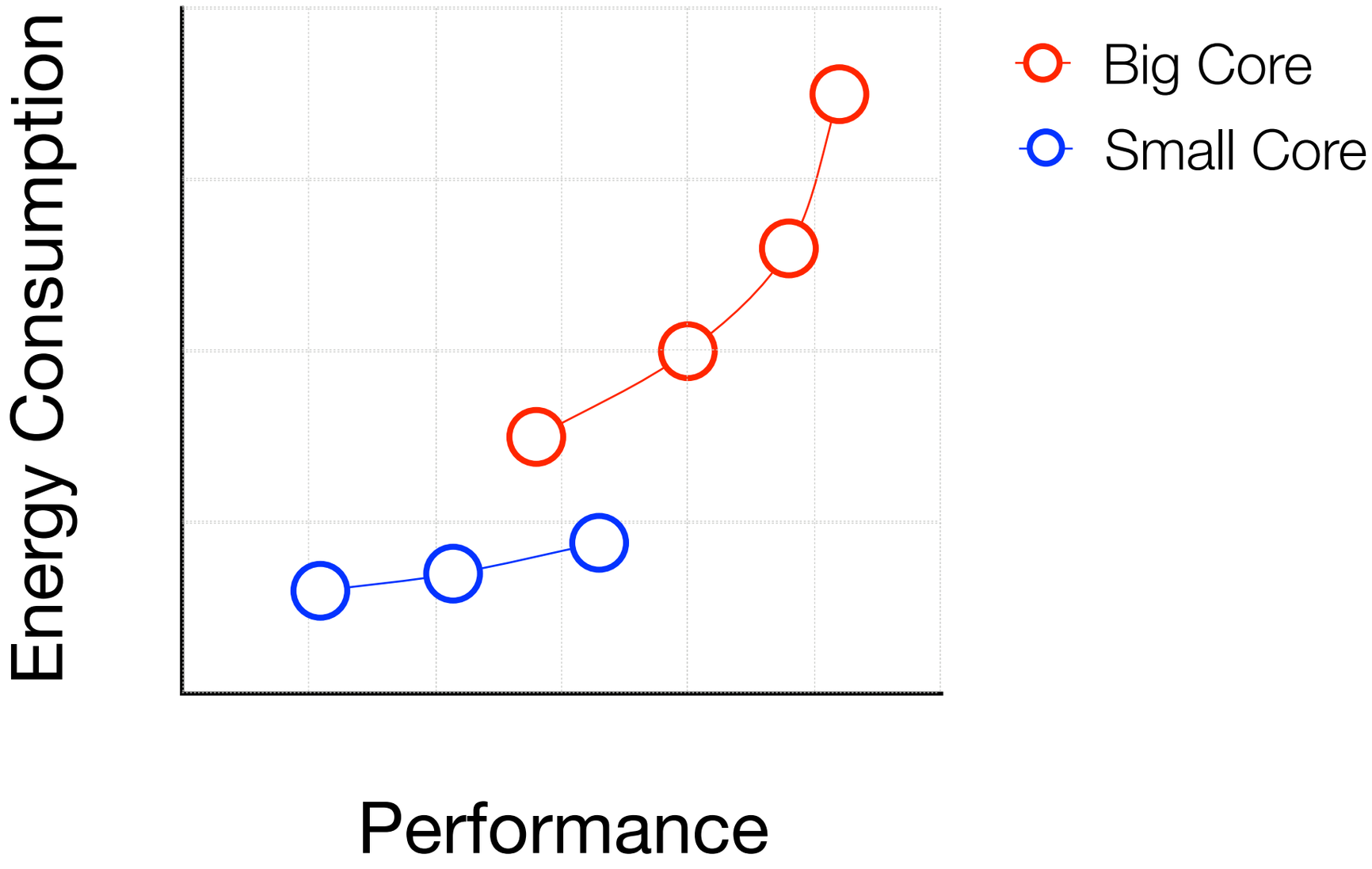
- Provide just enough energy to meet QoS constraints



ACMP-based **GreenWeb** Runtime

- Provide just enough energy to meet QoS constraints

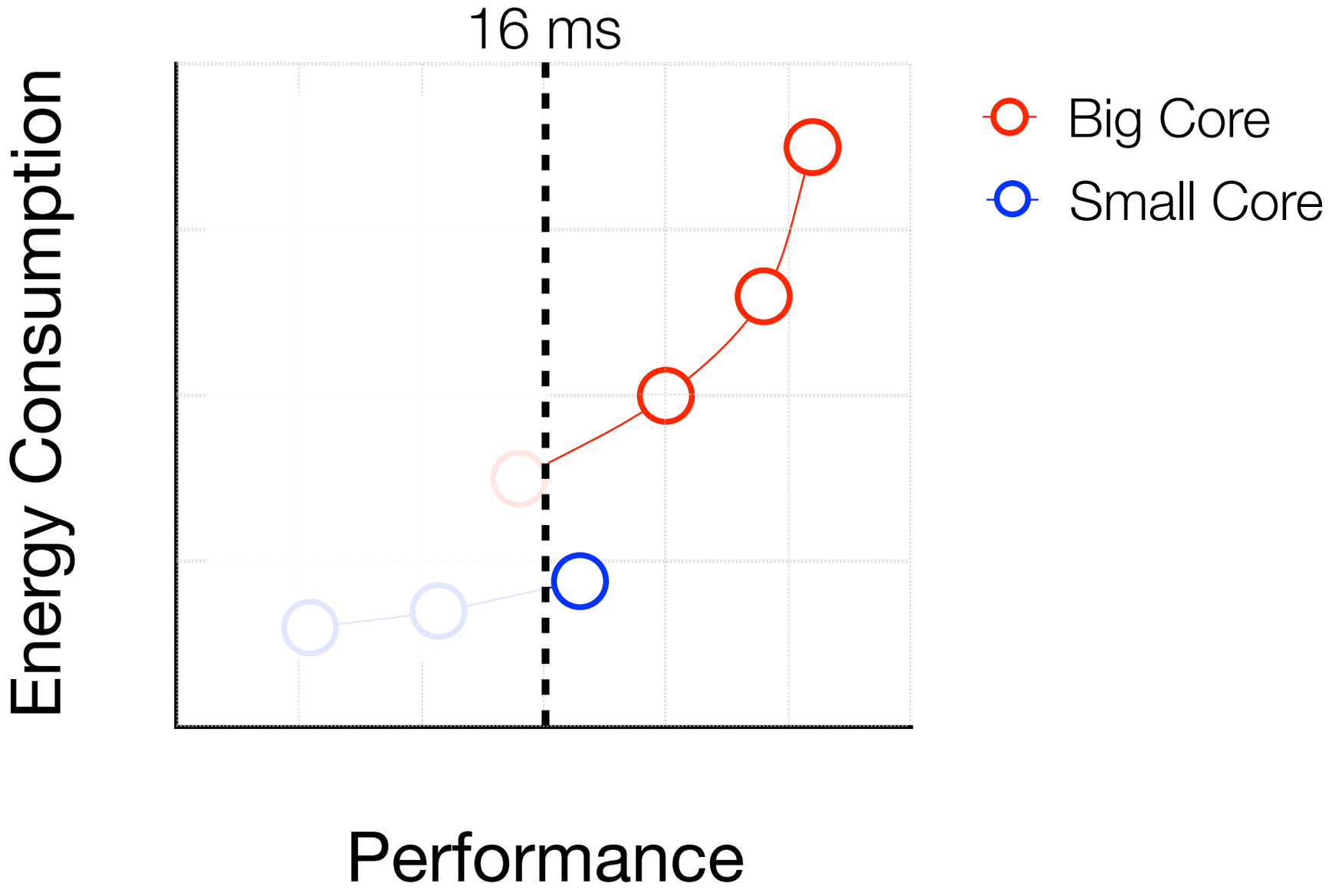
```
div {ontouchend: latency, 16 ms}
```



ACMP-based **GreenWeb** Runtime

- Provide just enough energy to meet QoS constraints

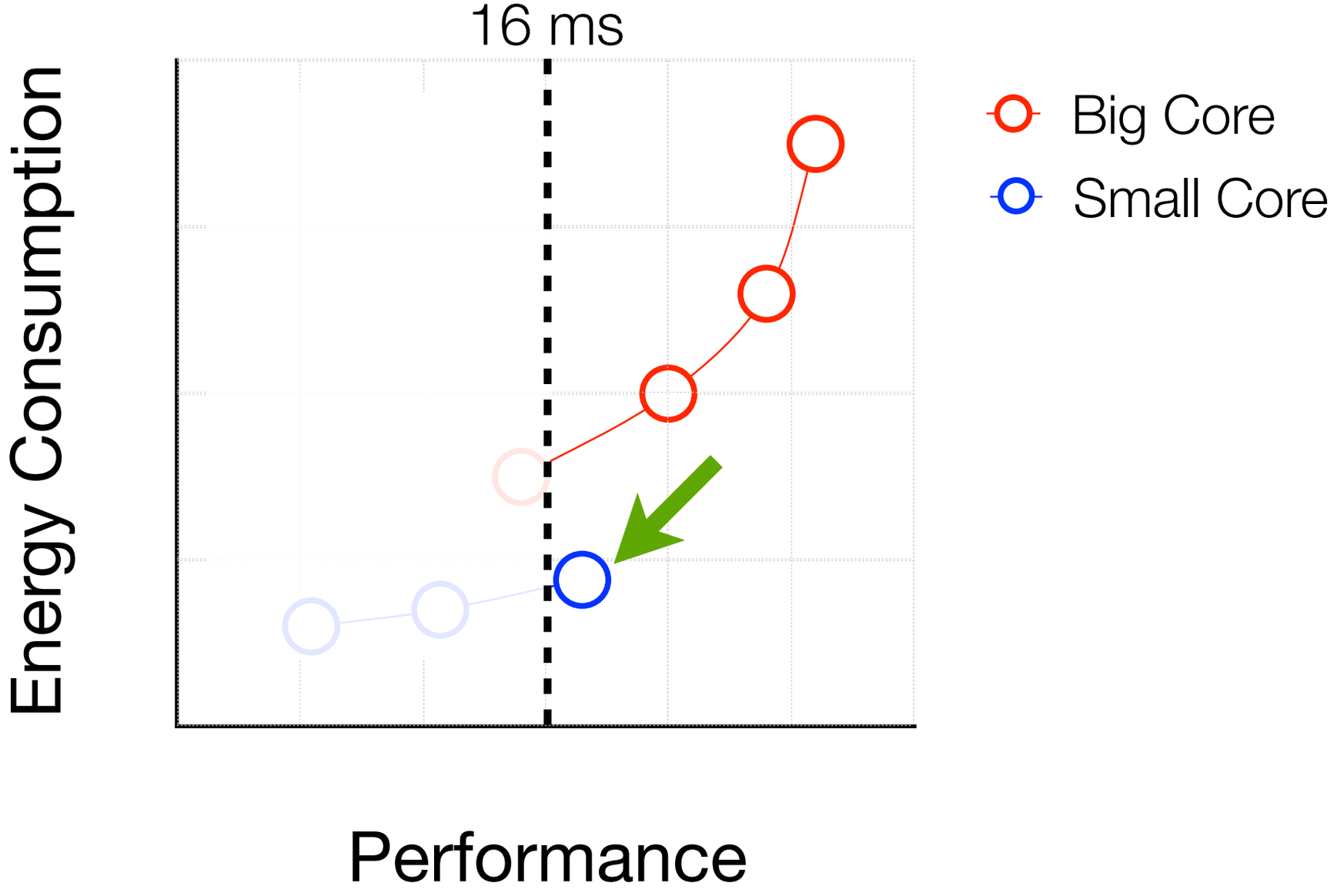
```
div {ontouchend: latency, 16 ms}
```



ACMP-based **GreenWeb** Runtime

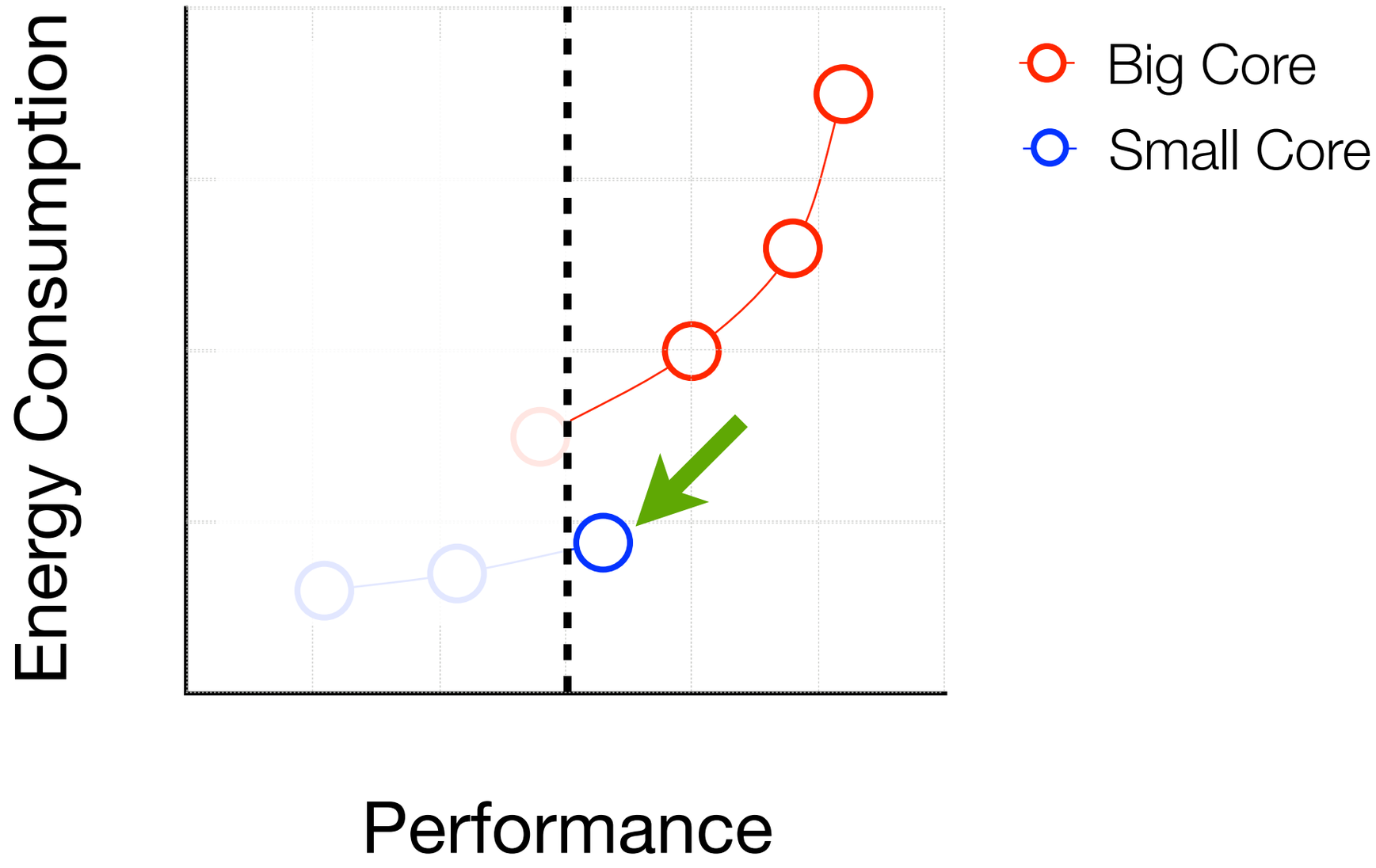
- Provide just enough energy to meet QoS constraints

```
div {ontouchend: latency, 16 ms}
```



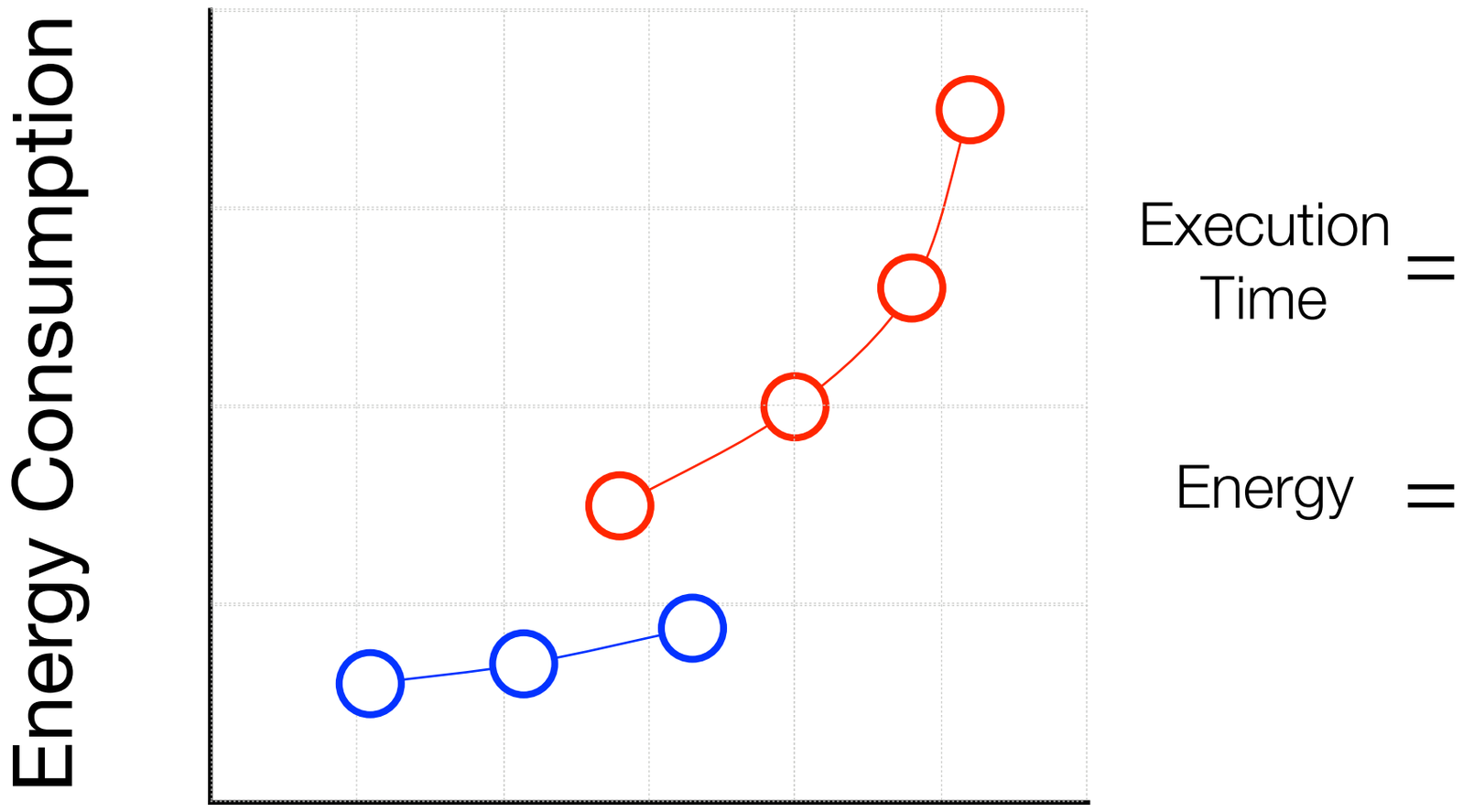
ACMP-based **GreenWeb** Runtime

- Provide just enough energy to meet QoS constraints
- Event-based scheduling [HPCA 2015]



ACMP-based **GreenWeb** Runtime

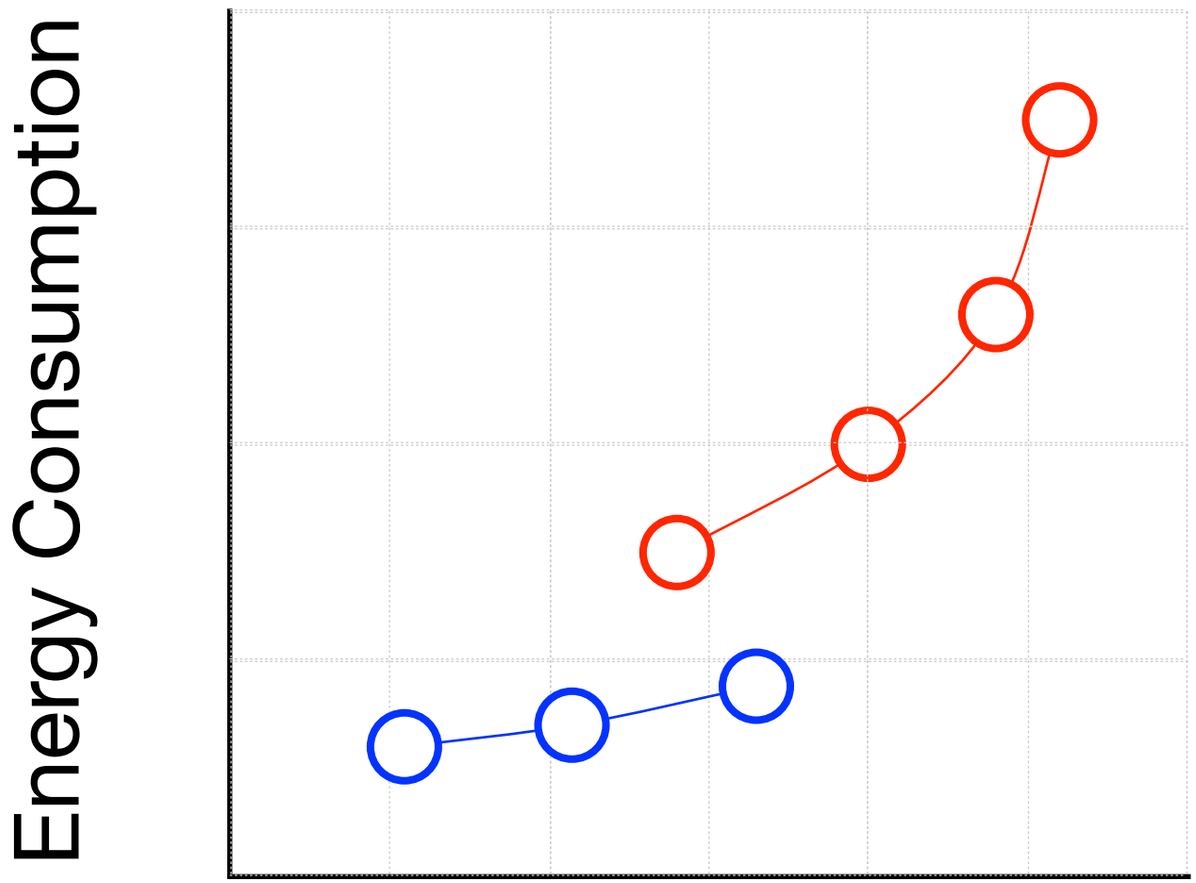
- Provide just enough energy to meet QoS constraints
- Event-based scheduling [HPCA 2015]



[PLDI 2003] Xie, et al., “Compile-Time Dynamic Voltage Scaling Settings: Opportunities and Limits”

ACMP-based **GreenWeb** Runtime

- Provide just enough energy to meet QoS constraints
- Event-based scheduling [HPCA 2015]



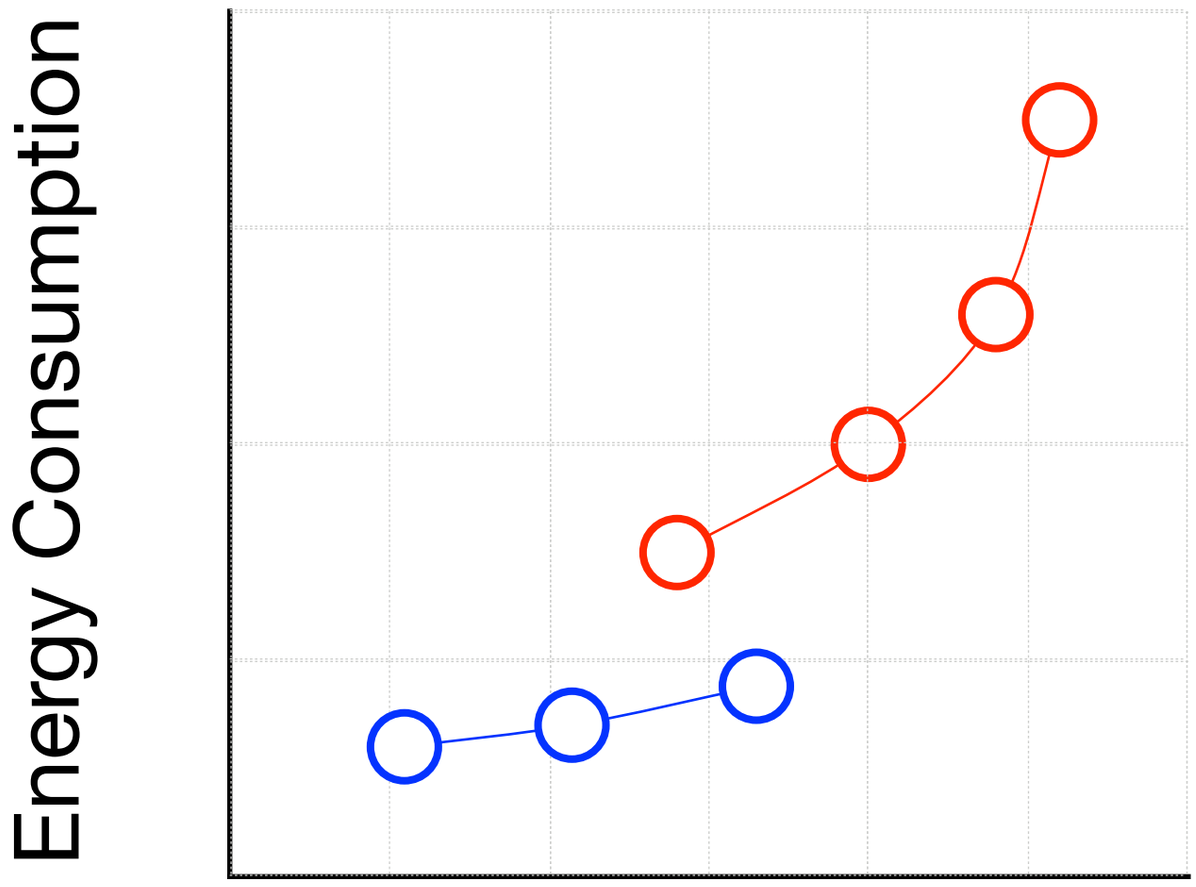
$$\text{Execution Time} = T_{memory} +$$

$$\text{Energy} =$$

[PLDI 2003] Xie, et al., “Compile-Time Dynamic Voltage Scaling Settings: Opportunities and Limits”

ACMP-based **GreenWeb** Runtime

- Provide just enough energy to meet QoS constraints
- Event-based scheduling [HPCA 2015]



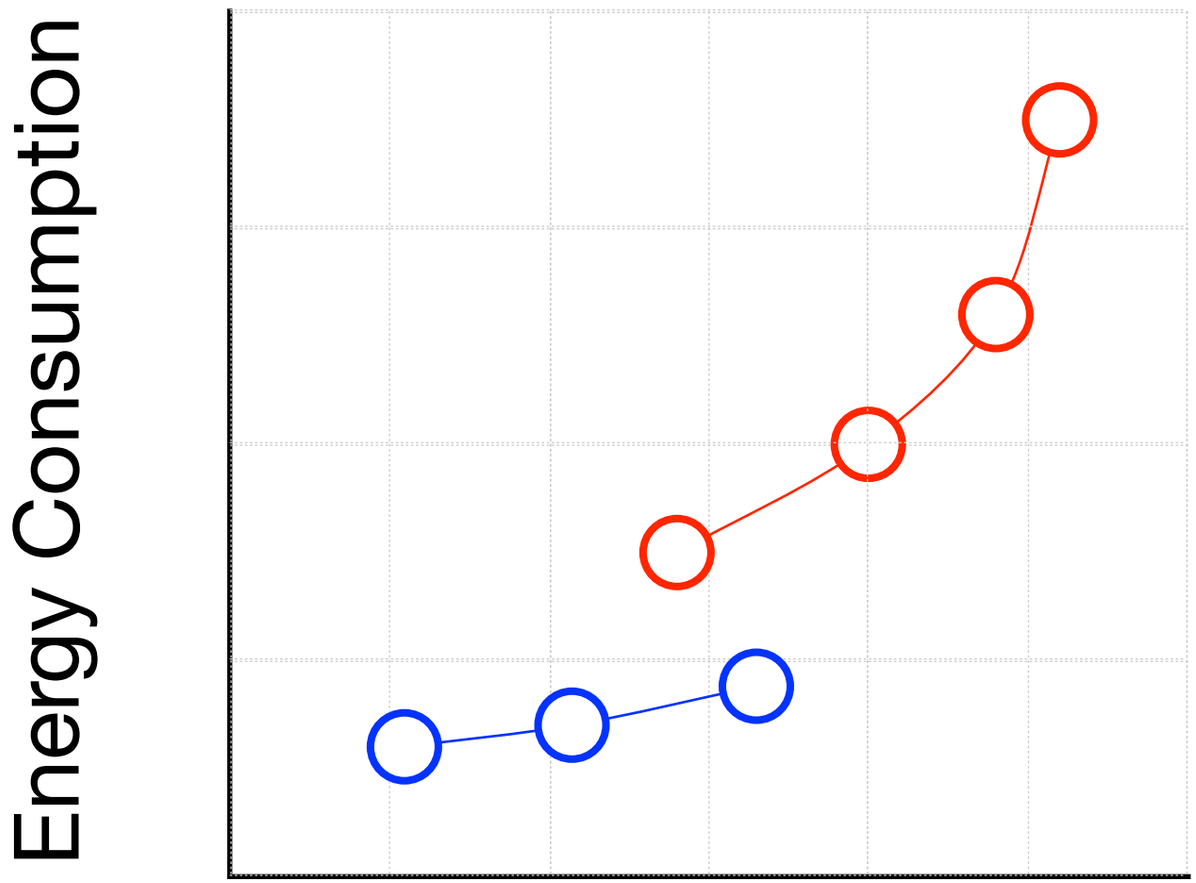
$$\text{Execution Time} = T_{memory} + T_{cpu}$$

$$\text{Energy} =$$

[PLDI 2003] Xie, et al., “Compile-Time Dynamic Voltage Scaling Settings: Opportunities and Limits”

ACMP-based **GreenWeb** Runtime

- Provide just enough energy to meet QoS constraints
- Event-based scheduling [HPCA 2015]



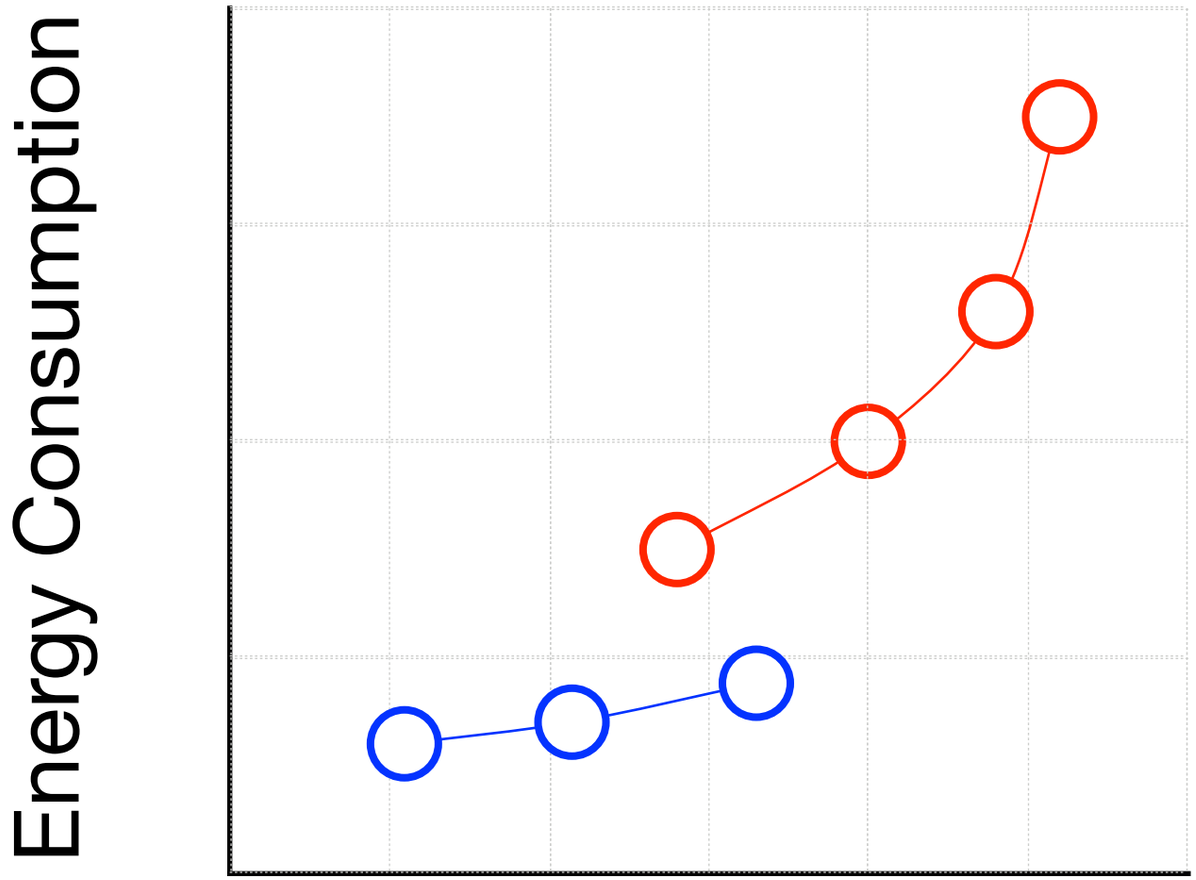
$$\text{Execution Time} = T_{memory} + N_{cycles} / f$$

$$\text{Energy} =$$

[PLDI 2003] Xie, et al., “Compile-Time Dynamic Voltage Scaling Settings: Opportunities and Limits”

ACMP-based **GreenWeb** Runtime

- Provide just enough energy to meet QoS constraints
- Event-based scheduling [HPCA 2015]



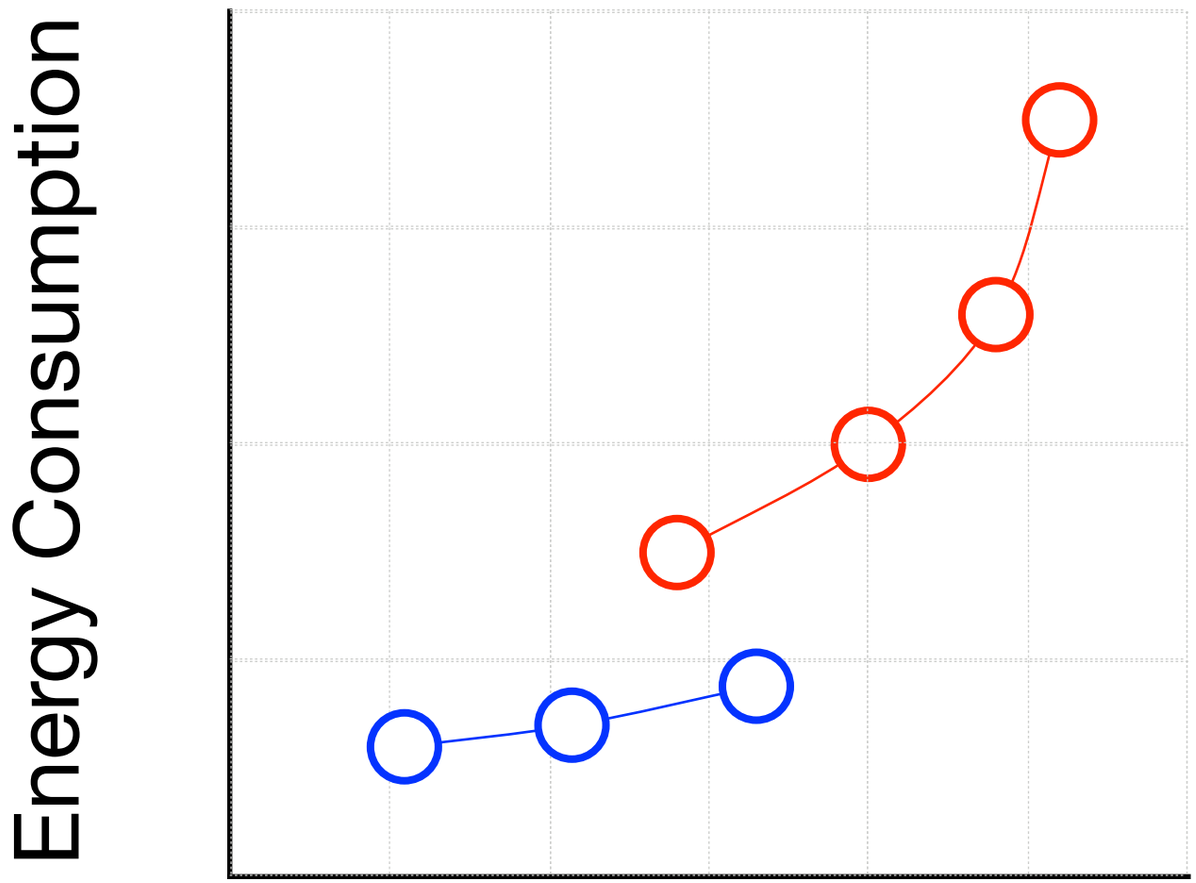
$$\text{Execution Time} = \boxed{T_{memory}} + \boxed{N_{cycles}} / f$$

$$\text{Energy} =$$

[PLDI 2003] Xie, et al., “Compile-Time Dynamic Voltage Scaling Settings: Opportunities and Limits”

ACMP-based **GreenWeb** Runtime

- Provide just enough energy to meet QoS constraints
- Event-based scheduling [HPCA 2015]

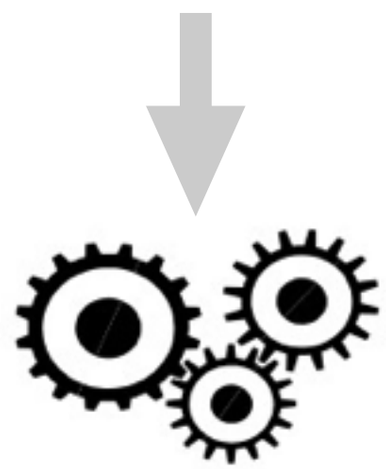


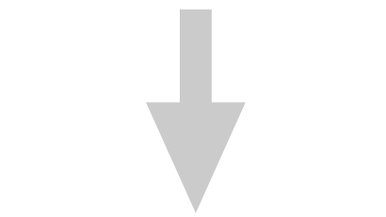
$$\text{Execution Time} = T_{memory} + N_{cycles} / f$$

$$\text{Energy} = \text{Execution Time} \times \text{Power}$$

[PLDI 2003] Xie, et al., “Compile-Time Dynamic Voltage Scaling Settings: Opportunities and Limits”

GreenWeb: Language for Energy-Efficiency





- Language abstractions
- **Runtime** that saves energy while meeting the QoS constraints
- Result
  hardware/software implementations

GreenWeb: Language for Energy-Efficiency

- Language abstractions

- Runtime
  the QoS constraints

- **Result** in 60% energy savings on real hardware/software implementations

Real Hardware/Software Setup

ODroid XU+E development board,
which contains an Exynos 5410 SoC
used in Samsung Galaxy S4.

Real Hardware/Software Setup

ODroid XU+E development board, which contains an Exynos 5410 SoC used in Samsung Galaxy S4.

Implementation incorporated into Chrome running on Android.

Real Hardware/Software Setup

ODroid XU+E development board, which contains an Exynos 5410 SoC used in Samsung Galaxy S4.

Implementation incorporated into Chrome running on Android.

MOSAIC

UI-level record and replay for reproducibility. [ISPASS'15]

Evaluation

- Baseline Mechanisms
  - Highest performance (**Perf**) — Standard to guarantee responsiveness
  - Interactive governor (**Interactive**) — Android default

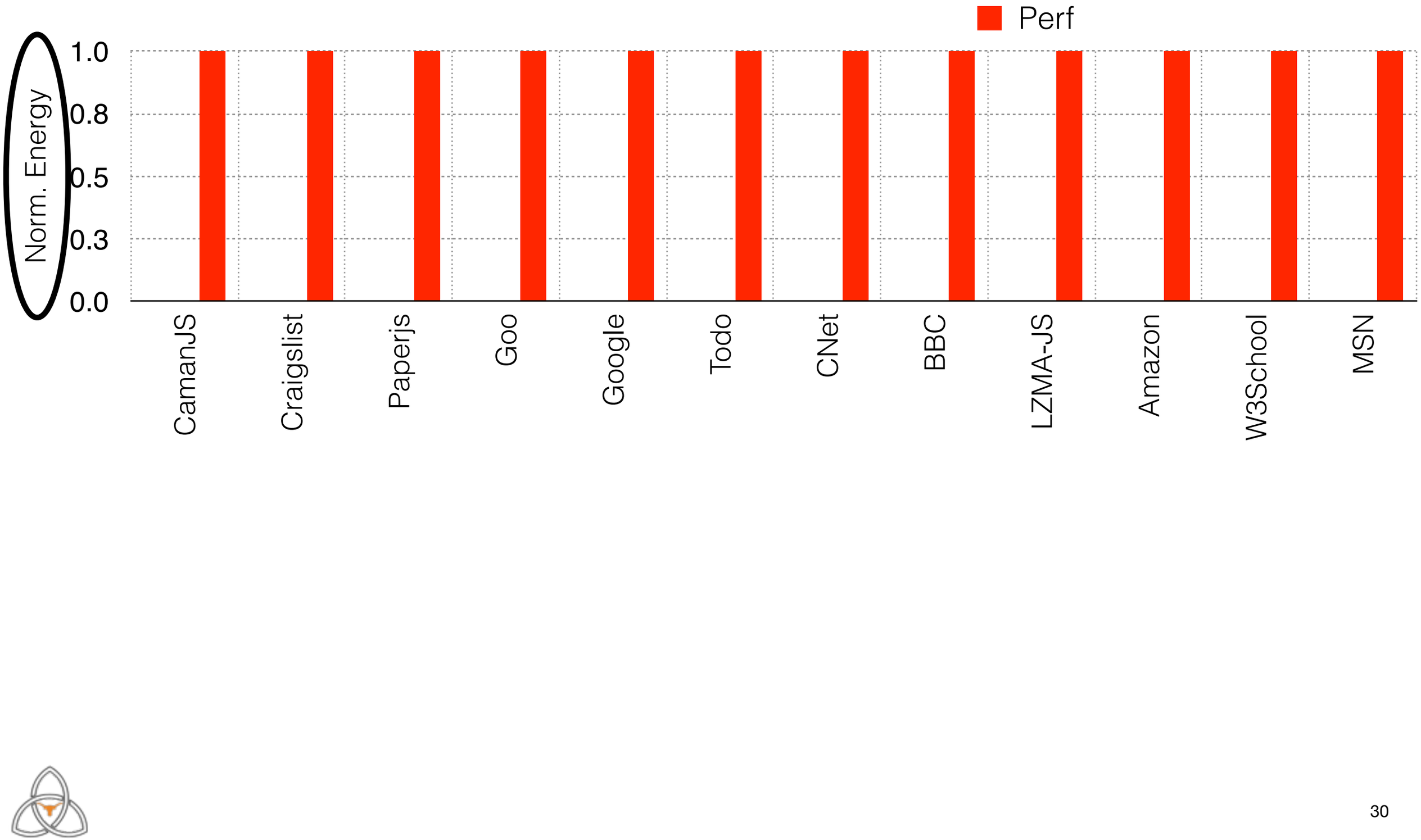
Evaluation

- Baseline Mechanisms
  - Highest performance (**Perf**) — Standard to guarantee responsiveness
  - Interactive governor (**Interactive**) — Android default
- Metrics
  - Energy Saving
  - QoS Violation

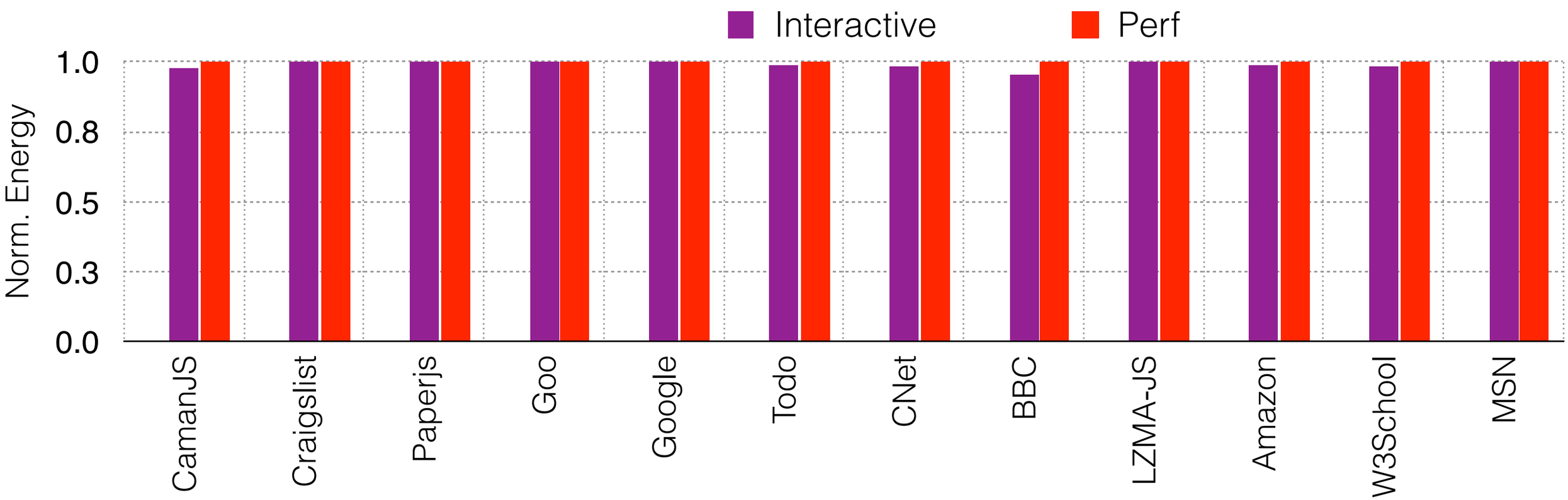
Evaluation

- Baseline Mechanisms
  - Highest performance (**Perf**) — Standard to guarantee responsiveness
  - Interactive governor (**Interactive**) — Android default
- Metrics
  - Energy Saving
  - QoS Violation
- Applications
  - Top webpages (e.g., www.amazon.com)
  - Web Apps based on popular frameworks (e.g., Todo List)

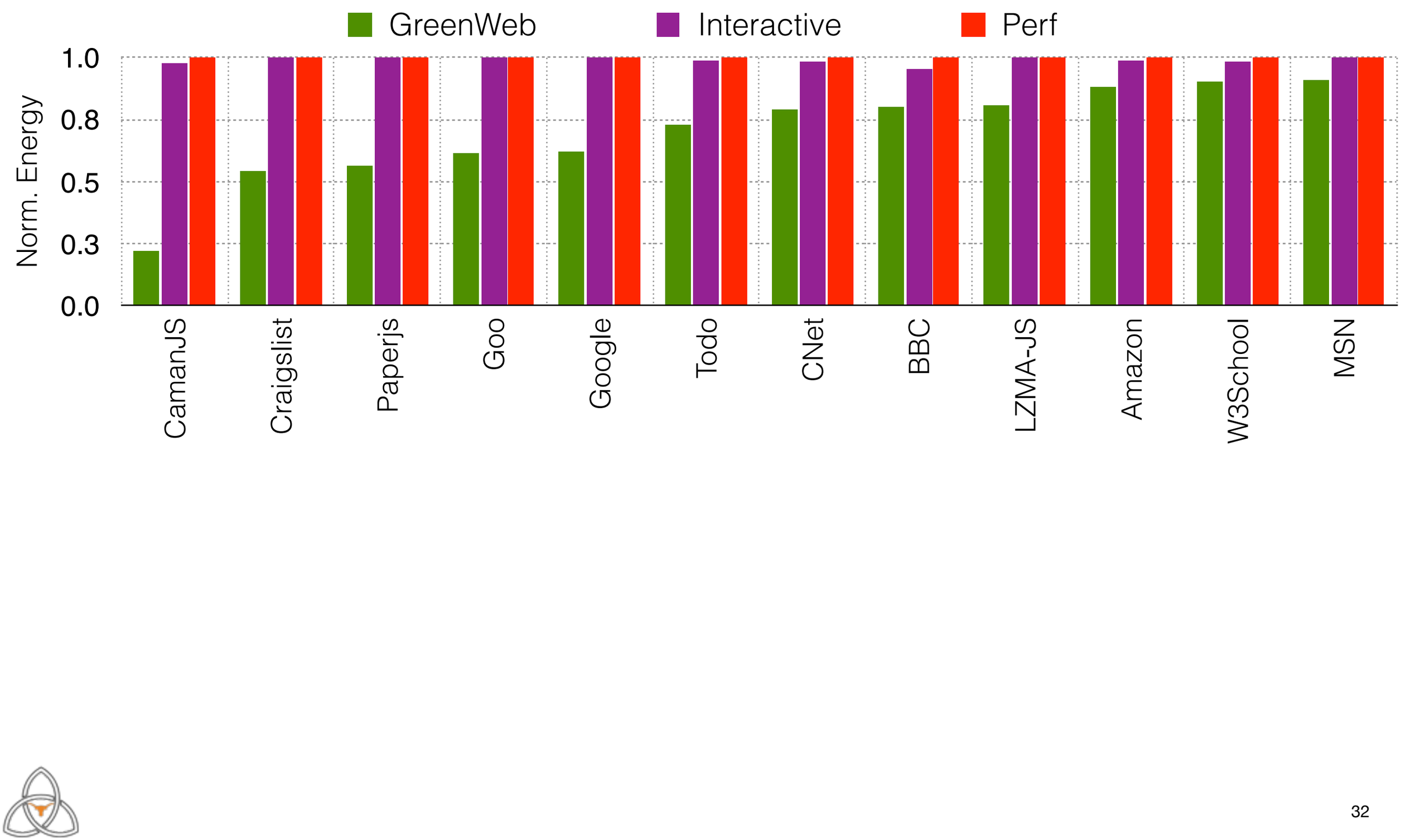
Evaluation Results



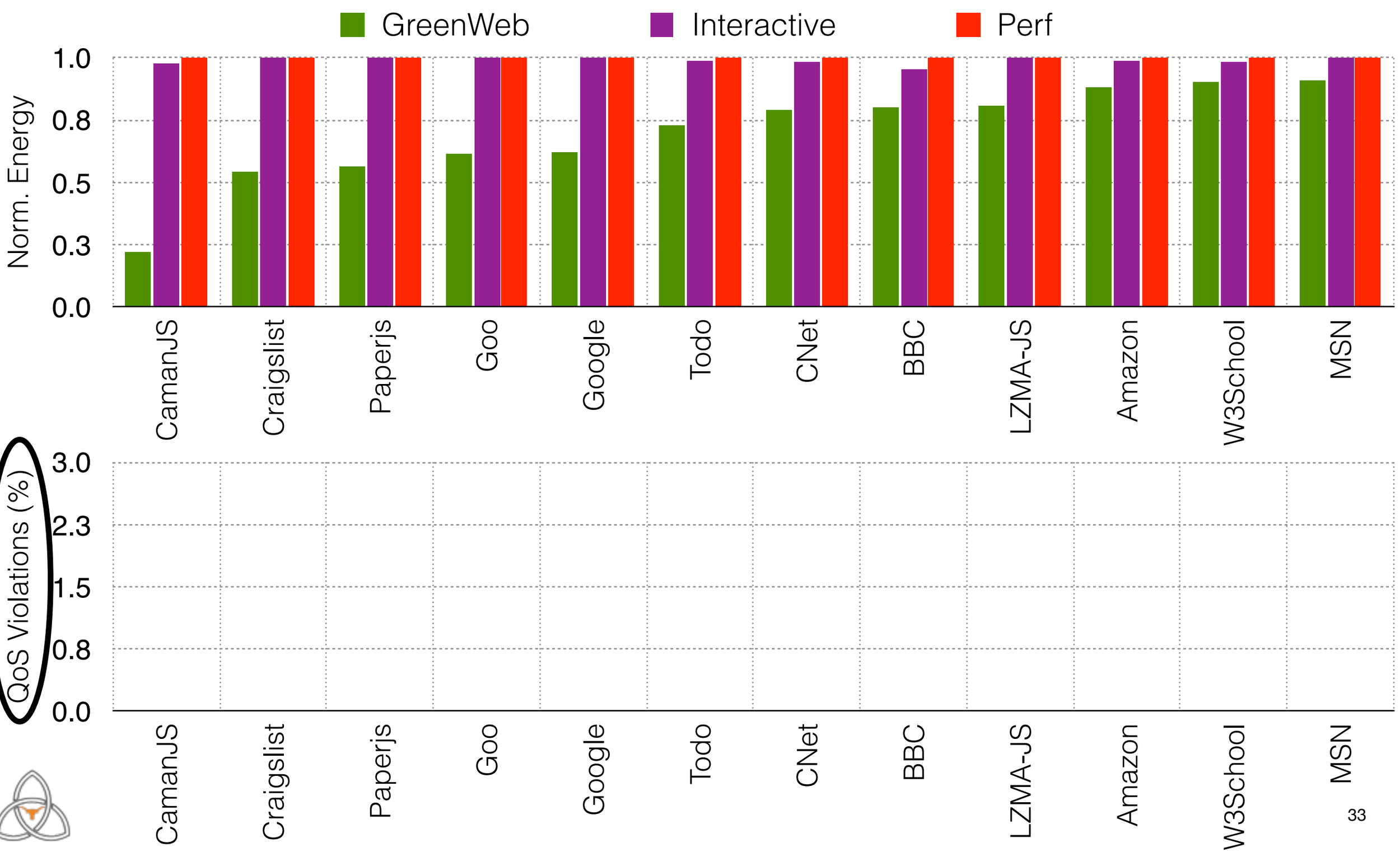
Evaluation Results



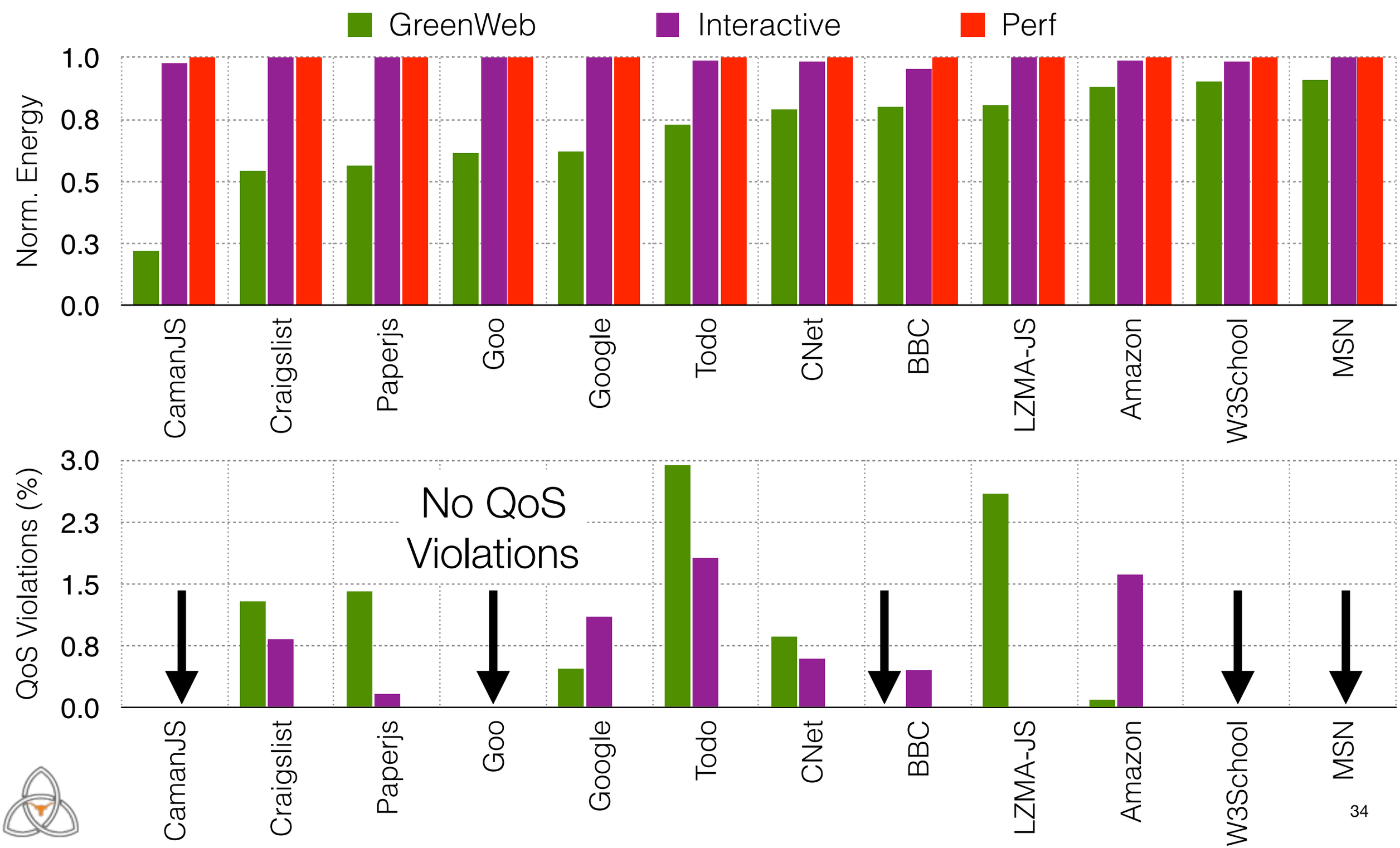
Evaluation Results



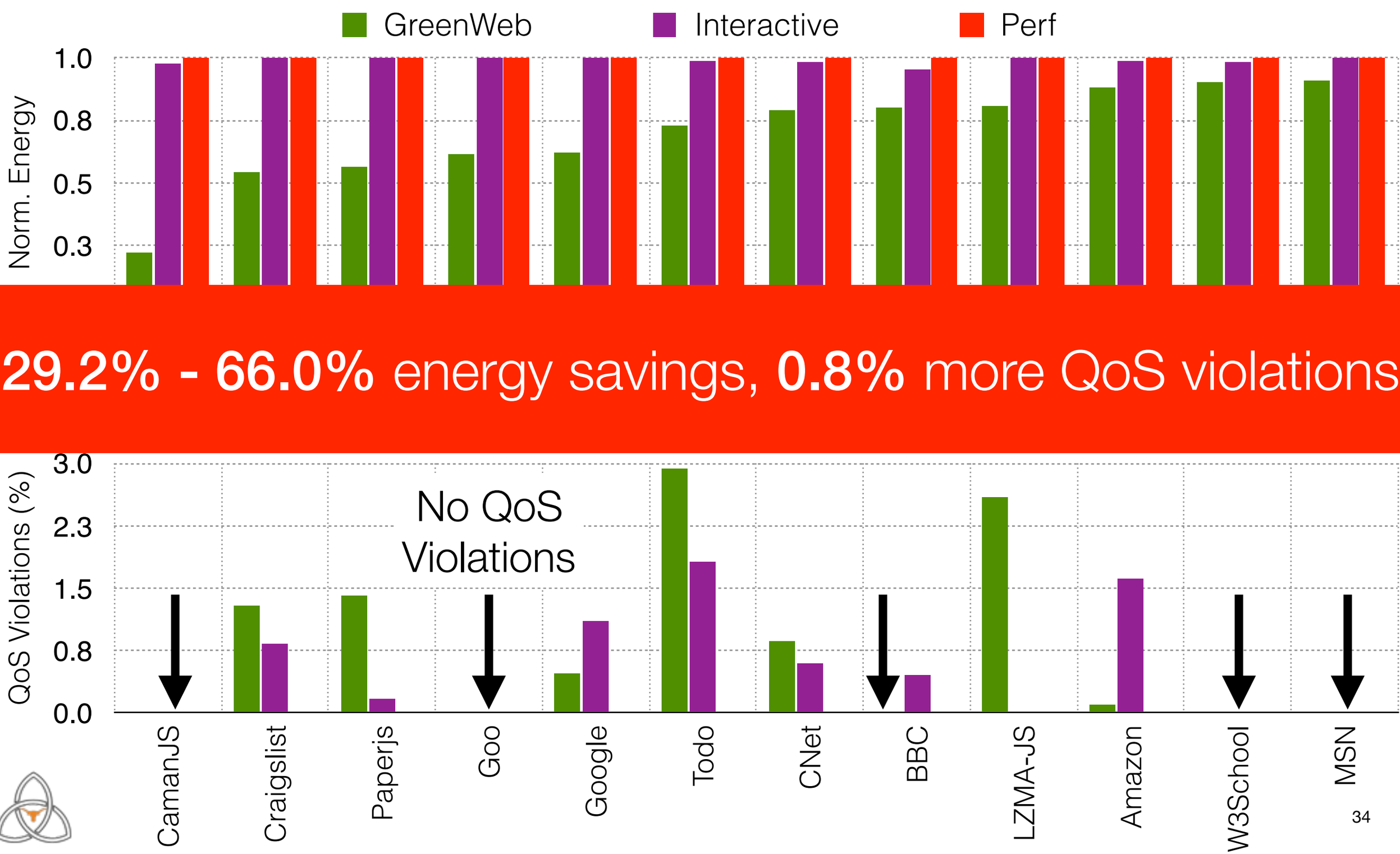
Evaluation Results



Evaluation Results



Evaluation Results



29.2% - 66.0% energy savings, **0.8%** more QoS violations

GreenWeb

Programming language support for balancing energy-efficiency and QoS in mobile Web computing

GreenWeb

Programming language support for balancing energy-efficiency and QoS in mobile Web computing

Abstraction

Express QoS constraints

GreenWeb

Programming language support for balancing energy-efficiency and QoS in mobile Web computing

Abstraction

Express QoS constraints

Runtime

Satisfy QoS specifications using energy saving techniques

GreenWeb

Programming language support for balancing energy-efficiency and QoS in mobile Web computing

Abstraction — Express QoS constraints

Runtime — Satisfy QoS specifications using energy saving techniques

Effect — Significant energy savings

wattwiseweb.org

GreenWeb:

Language Extensions for Energy-Efficient Mobile Web Computing

Yuhao Zhu
The University of Texas at Austin
with Vijay Janapa Reddi