Imperial College London

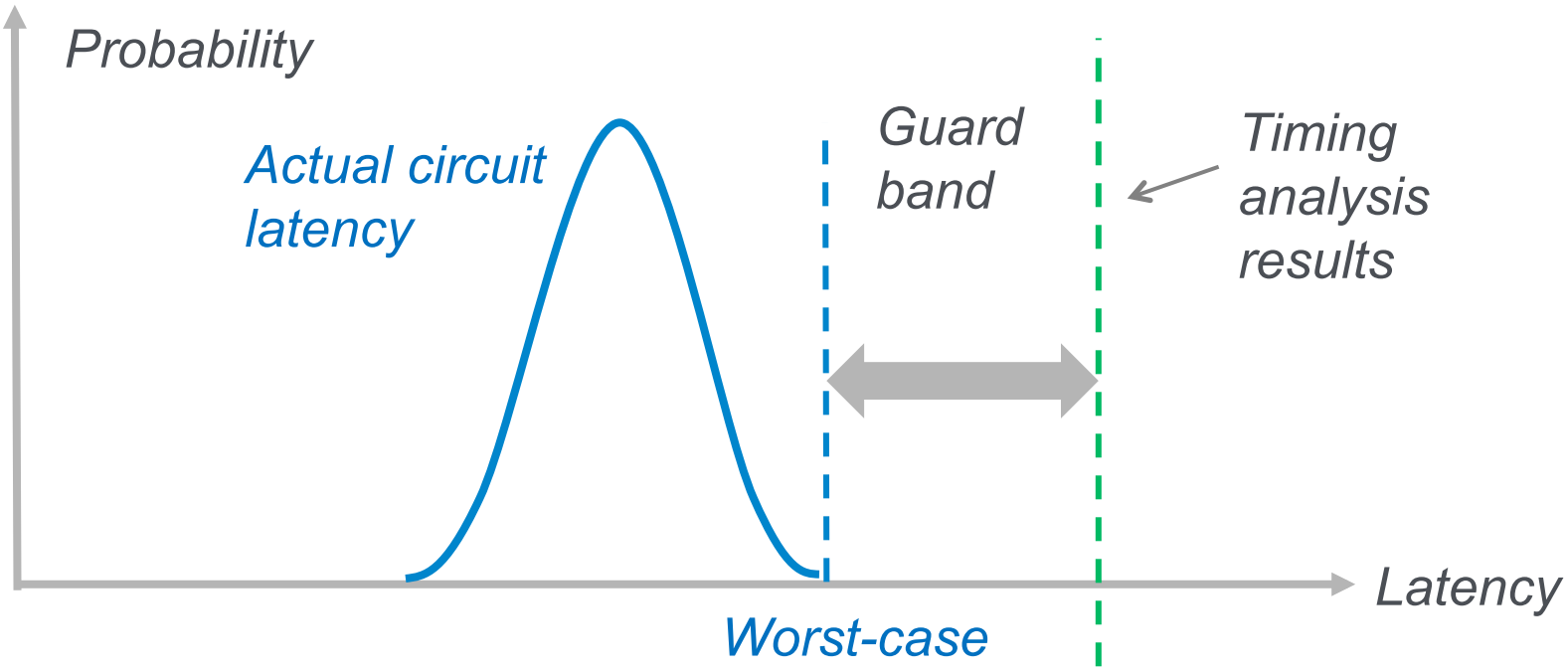
# Evaluation of Design Trade-offs for Adders in Approximate Datapath

**Kan Shi** and George A. Constantinides
EEE Department, Imperial College London

February 19, 2015
WAPCO'15 Amsterdam, Netherland

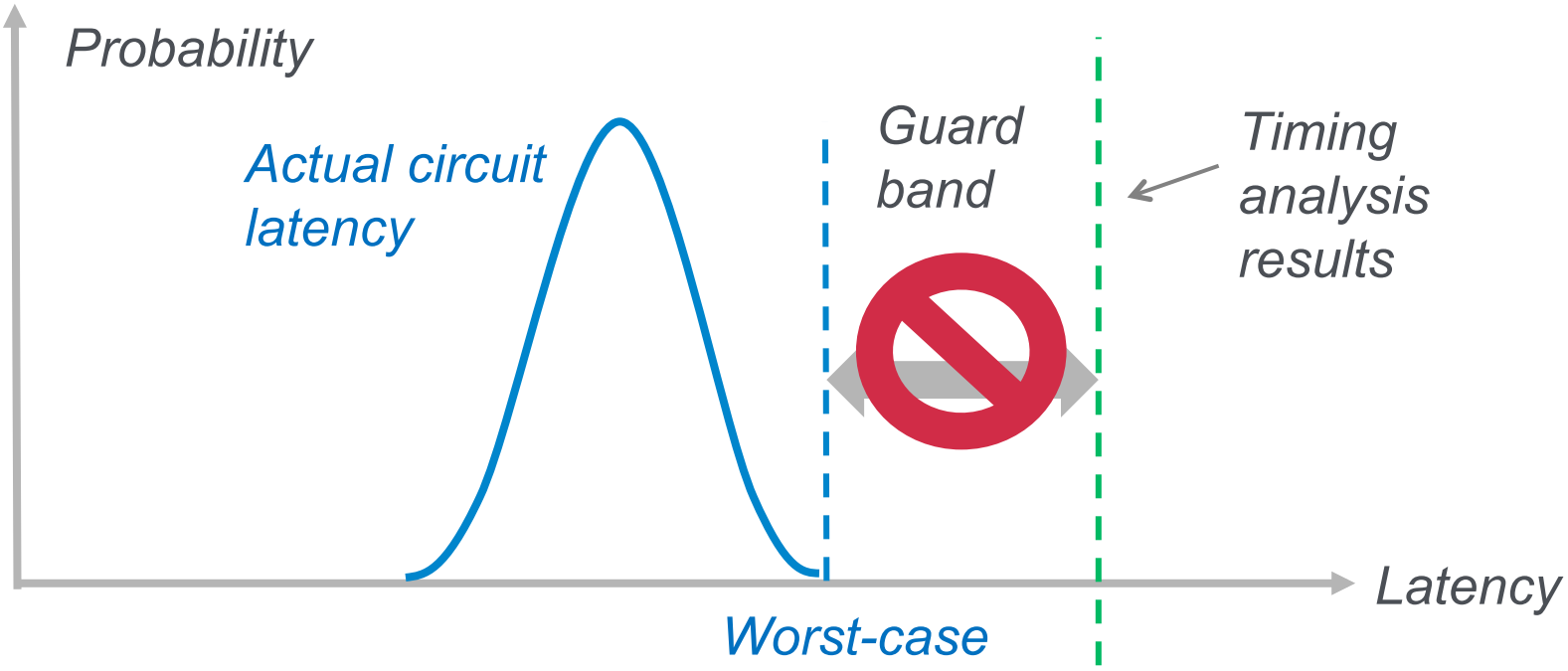
# It is all about Performance

- Digital circuits are designed to ensure timing closure



# It is all about Performance

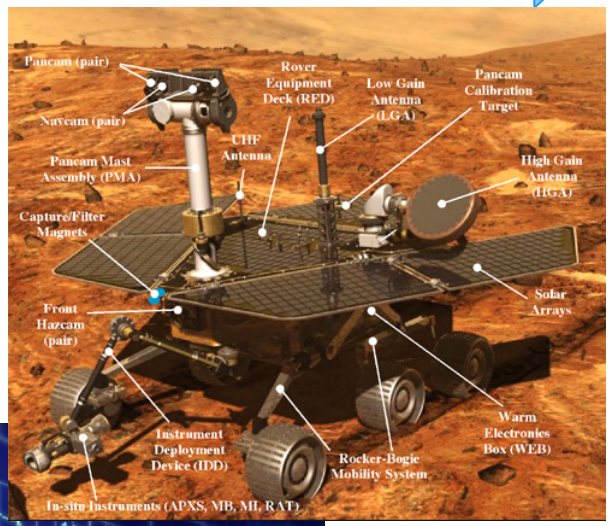
- Digital circuits are designed to ensure timing closure
- Designing for worst cases is Expensive
  - STA is conservative



Imperial College London

# Worst Case?

- Cost +



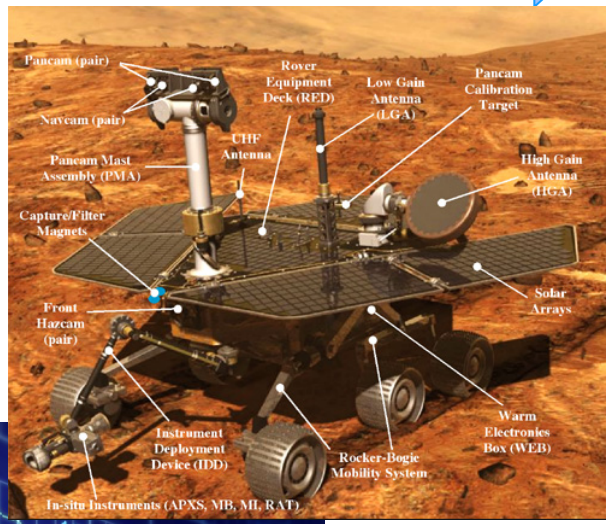


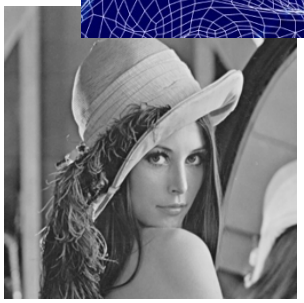
- Accuracy Requirement +

Imperial College London

# Worst Case?

- Cost +

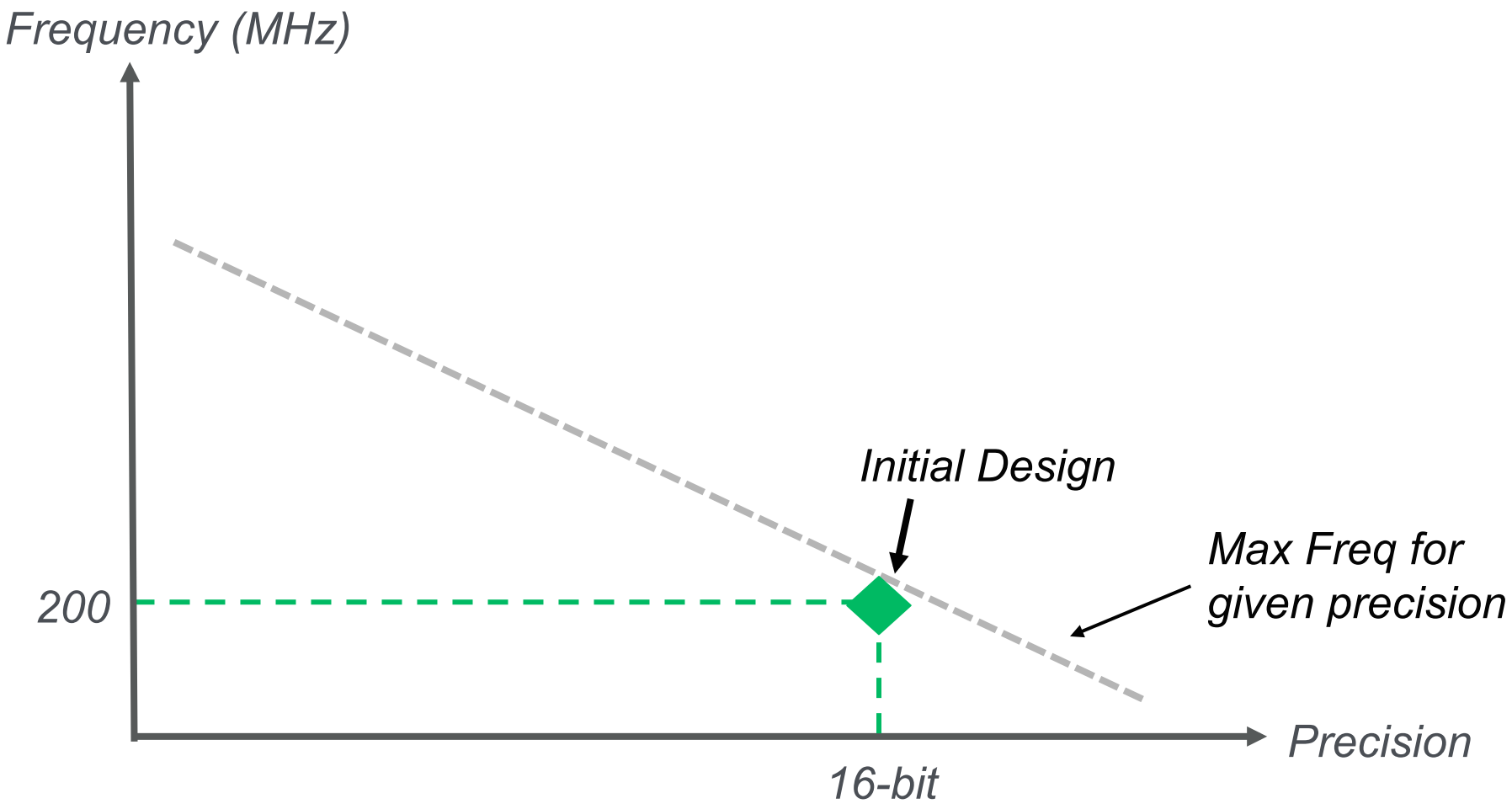



- Accuracy Requirement +

# Our Vision: Design Scenarios

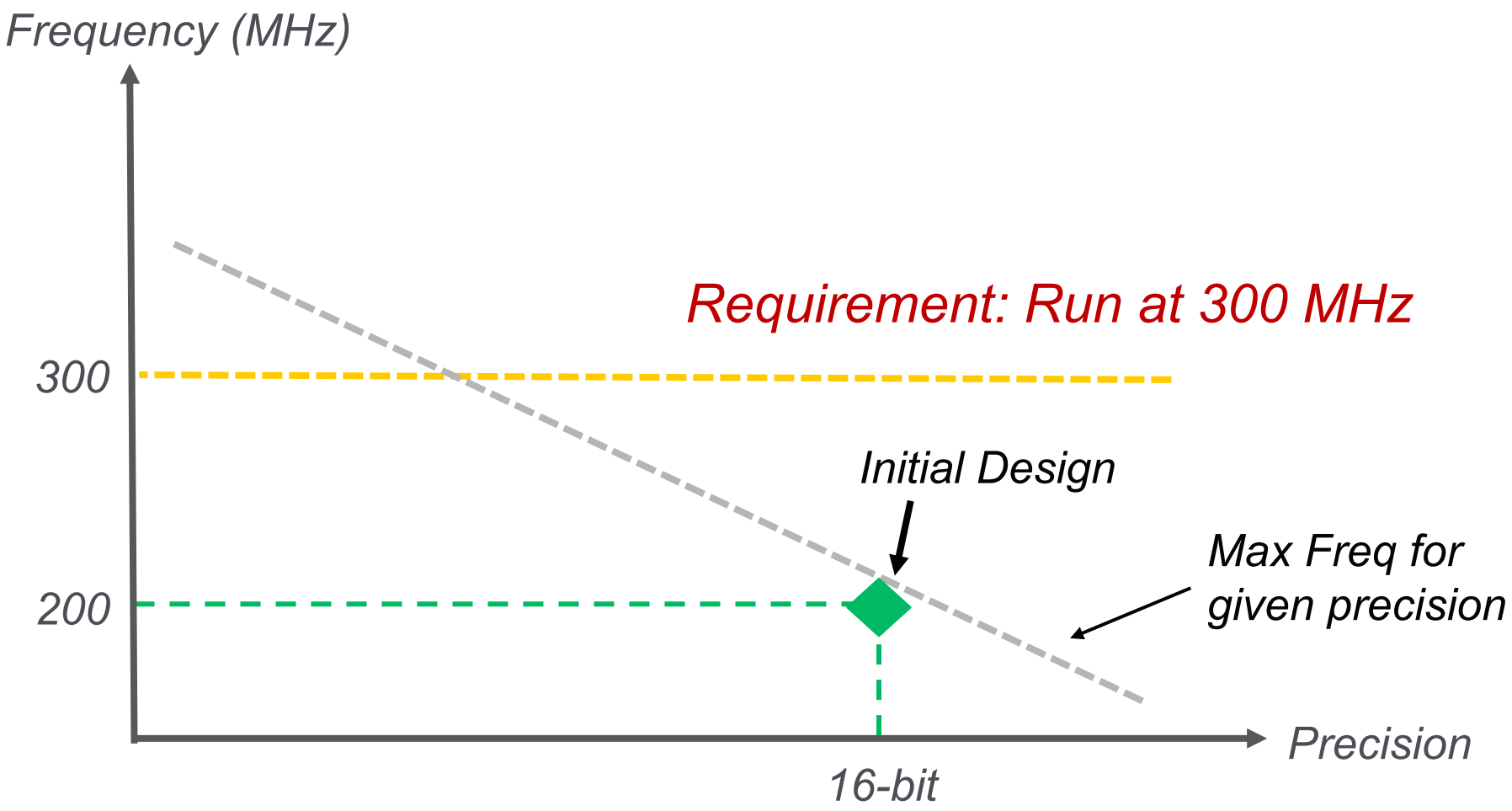
*(FCCM'13, ISCAS'13, TRETS'14)*



Imperial College London

# Our Vision: Design Scenarios

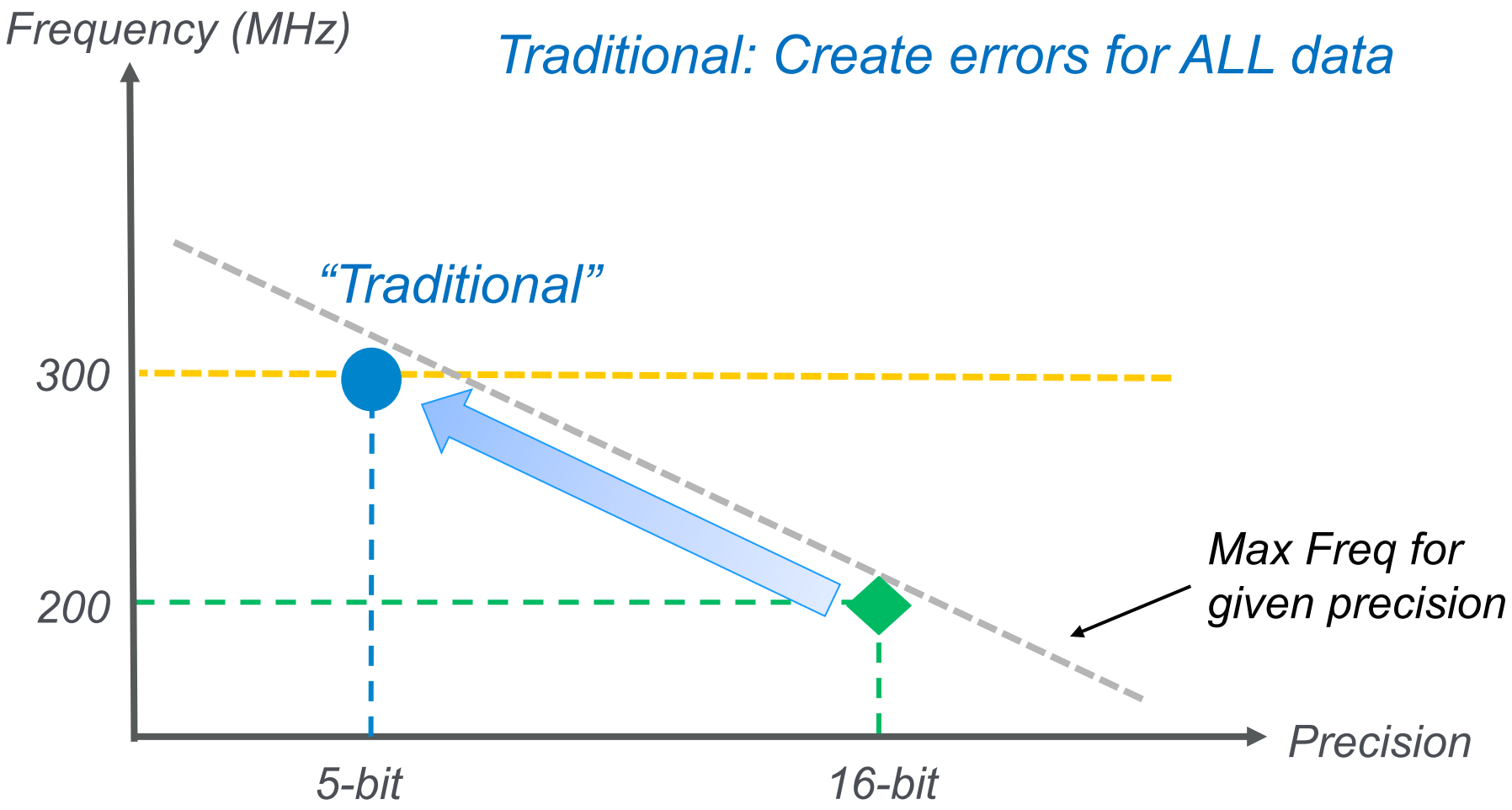
*(FCCM'13, ISCAS'13, TRETS'14)*



Imperial College London

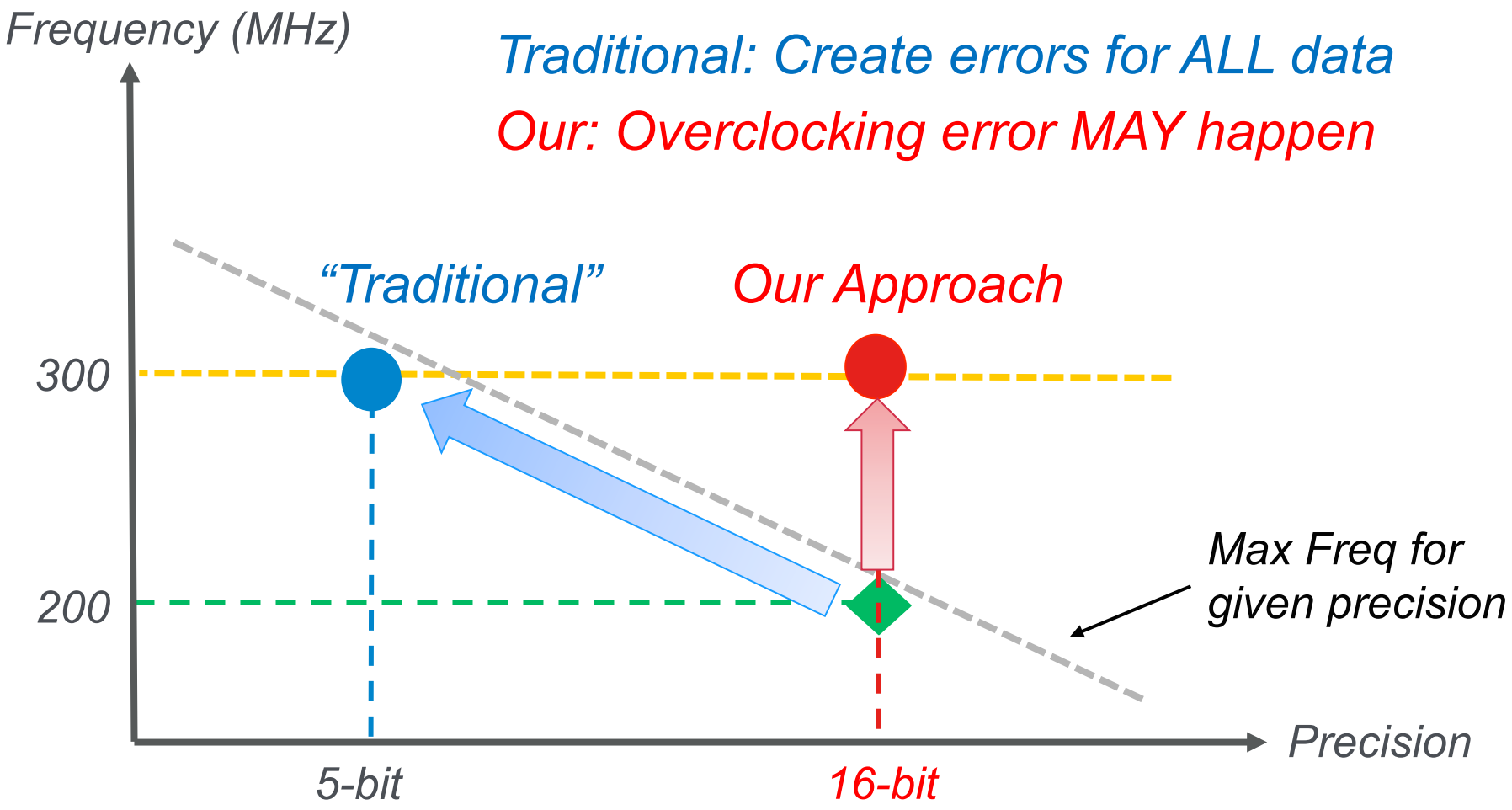
# Our Vision: Design Scenarios

*(FCCM'13, ISCAS'13, TRETS'14)*



# Our Vision: Design Scenarios

*(FCCM'13, ISCAS'13, TRETS'14)*



# Our Vision: Different Computer Arithmetic

*(DAC'14, ICFPT'14)*

- "Online Arithmetic" is MSB-first arithmetic
  - All I/Os are processed in the MSB-first manner
- Online Arithmetic is "overclocking friendly"
  - Overclocking errors only occur at LSBs.

# Our Vision: Different Computer Arithmetic

*(DAC'14, ICFPT'14)*

- "Online Arithmetic" is MSB-first arithmetic
  - All I/Os are processed in the MSB-first manner
- Online Arithmetic is "overclocking friendly"
  - Overclocking errors only occur at LSBs.

***Example: Image Filter (DAC'14), 25% Speed-up***

*Traditional Arithmetic*

*SNR=9dB*

*Online Arithmetic*

*SNR=36.2dB*

# Our Contributions: This Paper

- Detailed evaluation of three adder structures:
  - Ripple carry adder
  - Carry select adder
  - Online arithmetic adder

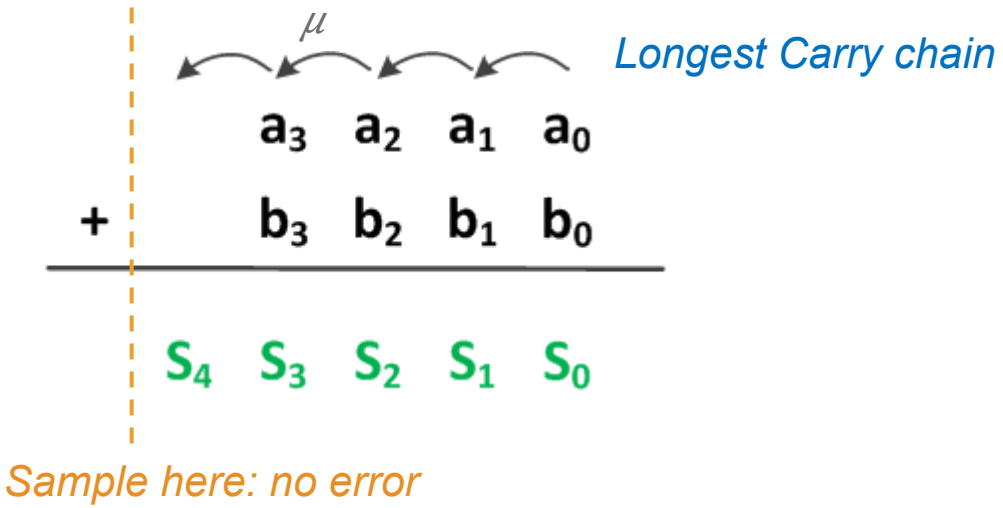
- Optimum design choice under accuracy, performance and area trade-offs

# Outline

- Motivation
- **Background**
  - **Adders with conventional arithmetic**
  - **Online arithmetic**
- Design trade-offs for adders
- Evaluation of the optimum adder design choice
- Conclusion

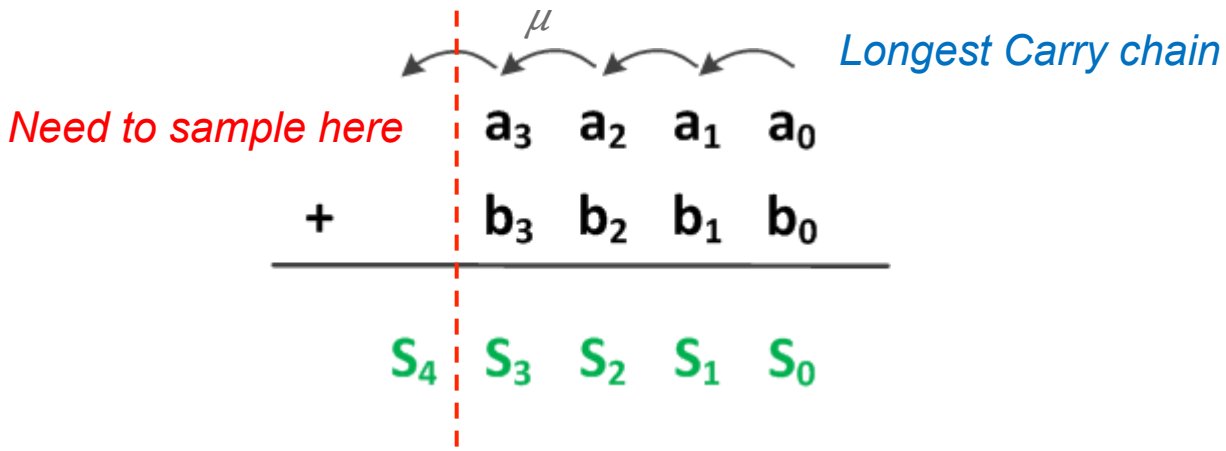
# Adders with Conventional Arithmetic

- Ripple carry adder
  - Frequency determines max carry propagation length



# Adders with Conventional Arithmetic

- Ripple carry adder
  - Frequency determines max carry propagation length



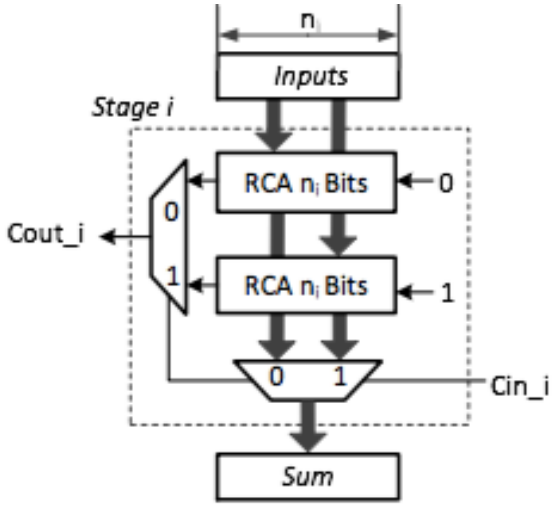

  - Two design scenarios: truncation or overclocking (see paper)

Imperial College London

# Adders with Conventional Arithmetic

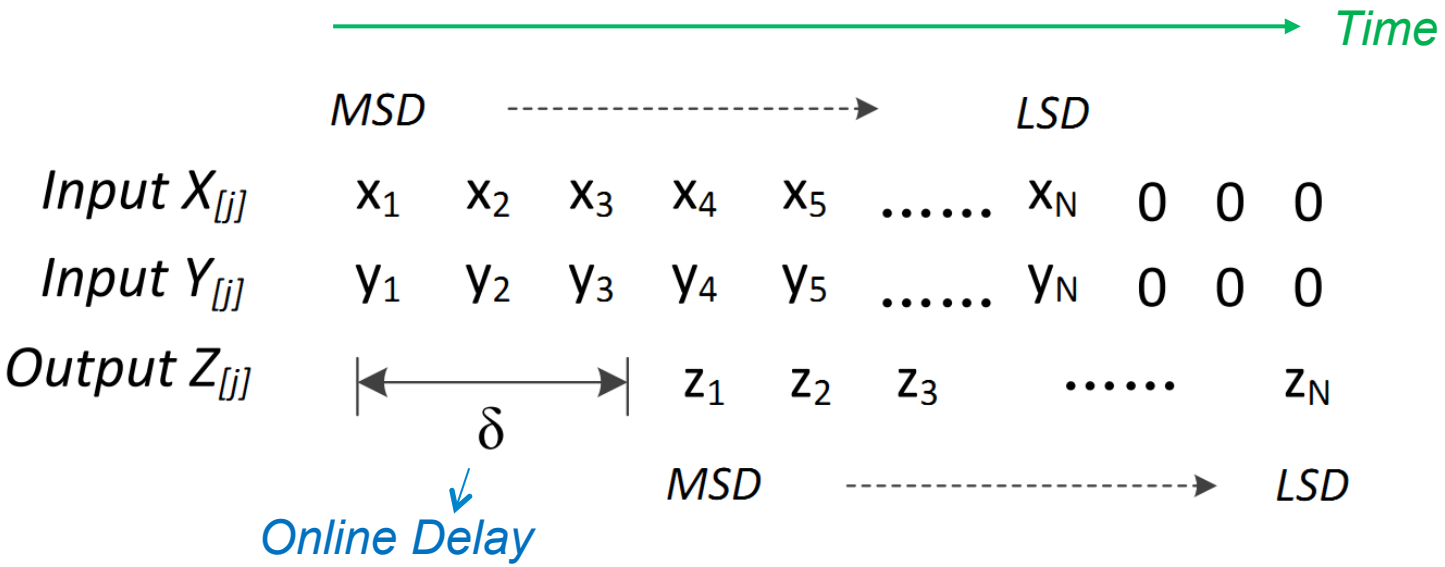
- Ripple carry adder
  - Frequency determines max carry propagation length
  - Two design scenarios: truncation or overclocking (see paper)

- Carry select adder
  - Use more area for high performance
  - Evaluate its truncation scenario
  - Details please see paper



# Online Arithmetic: Background

- What is it?
  - Online Arithmetic performs computation in an MSD-first manner.
  - It is initially designed for digit-serial operation.

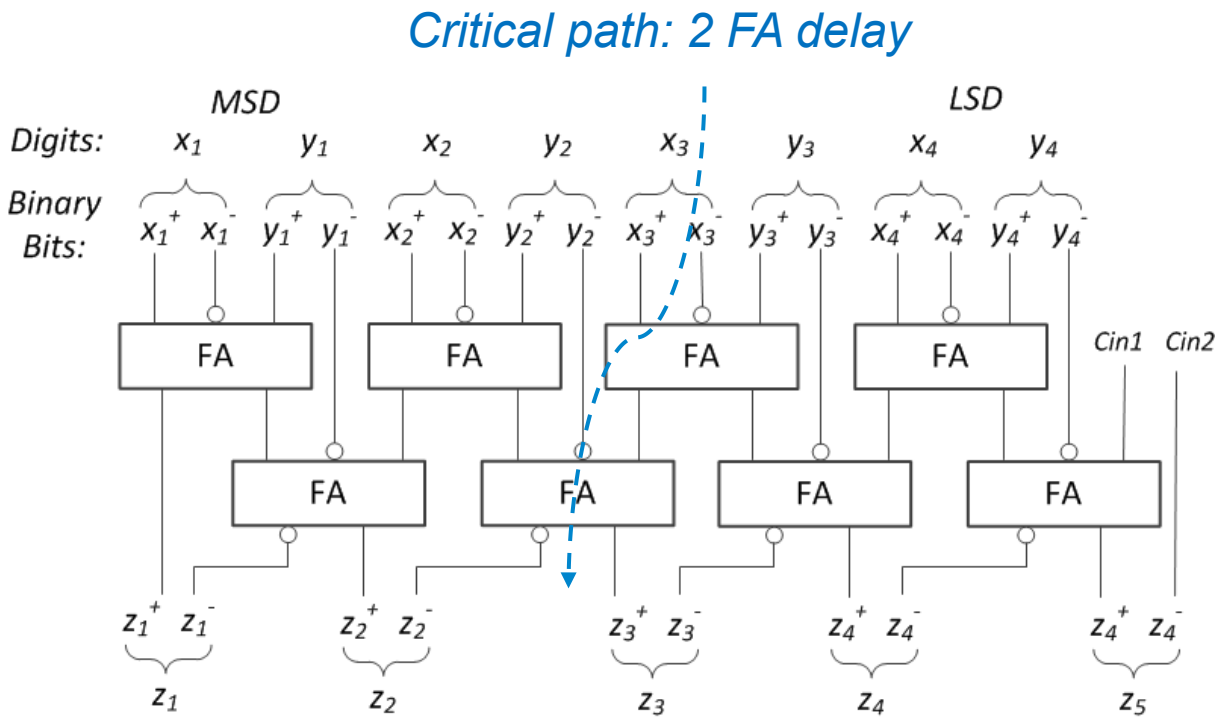


# Online Arithmetic: Background

- What is it?
  - Online Arithmetic performs computation in an **MSD-first manner**.
  - It is initially designed for digit-serial operation.
- How is this possible?
  - Requires a flexibility in computing outputs only based on partial information of inputs.
  - Redundancy in the number representation.

# Online Arithmetic: Adder

- Features:
  - Binary, digit parallel
  - Fast: critical path is irrelevant to operand precision

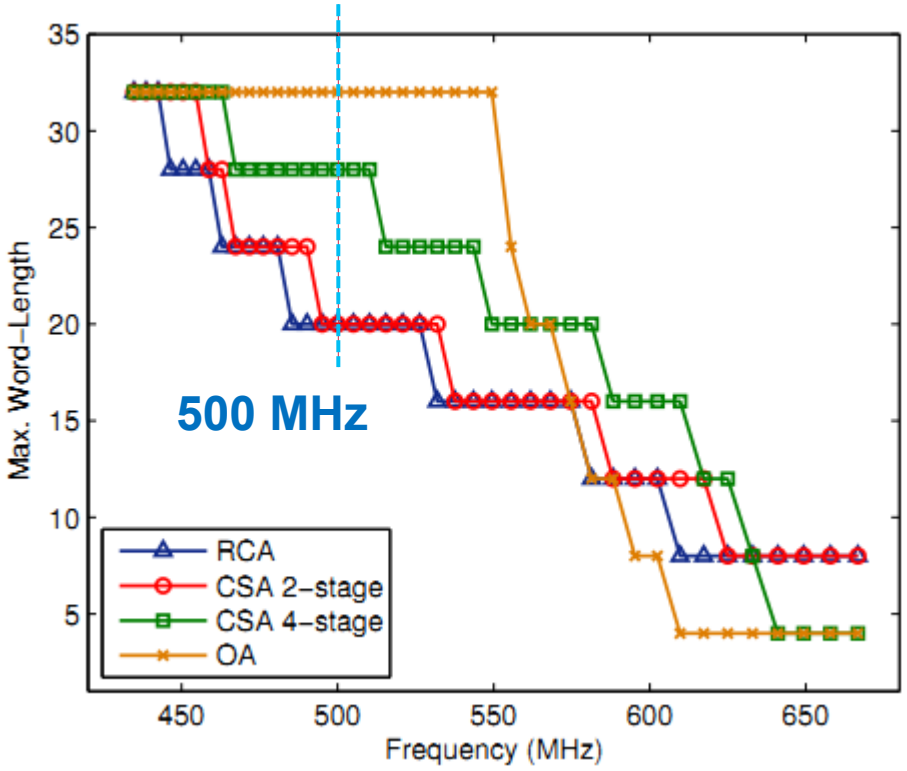


# Outline

- Motivation
- Background
- **Design trade-offs for adders**
  - **Accuracy, performance and area**
- Evaluation of the optimum adder design choice
- Conclusion

# Design Trade-offs: Accuracy vs Frequency

- Original word-length = 32 bits



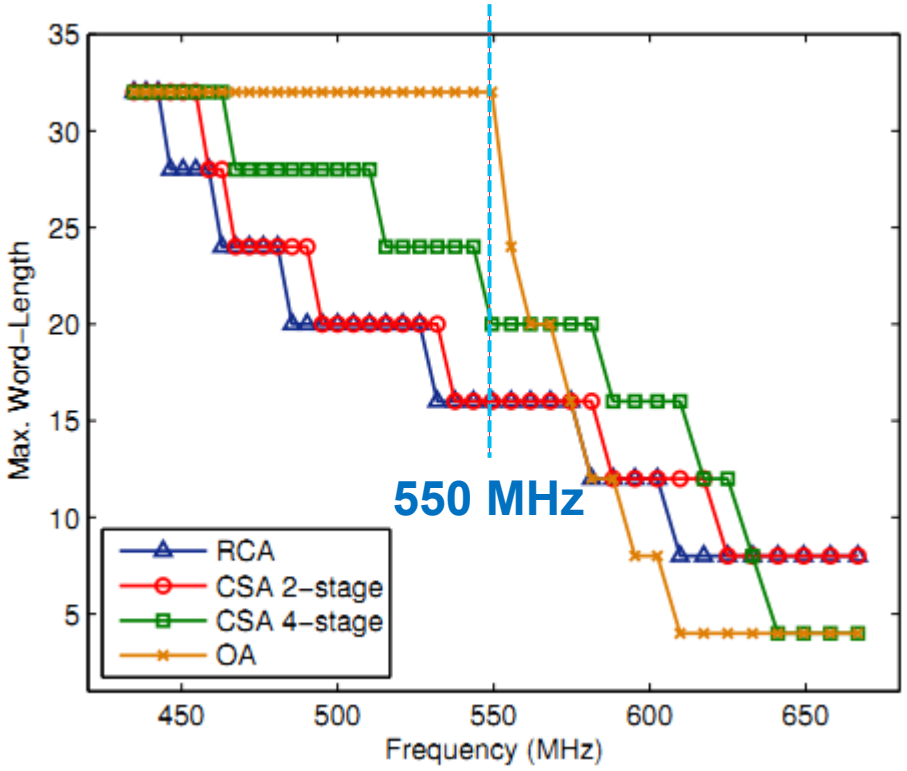

**Max. Word-length**

| Freq (MHz) | RCA | CSA-2 | CSA-4 | Online Adder |
|---|---|---|---|---|
| 500 | 20 | 20 | 28 | 32 |
| | | | | |
| | | | | |

Imperial College London

# Design Trade-offs: Accuracy vs Frequency

- Original word-length = 32 bits




**Max. Word-length**

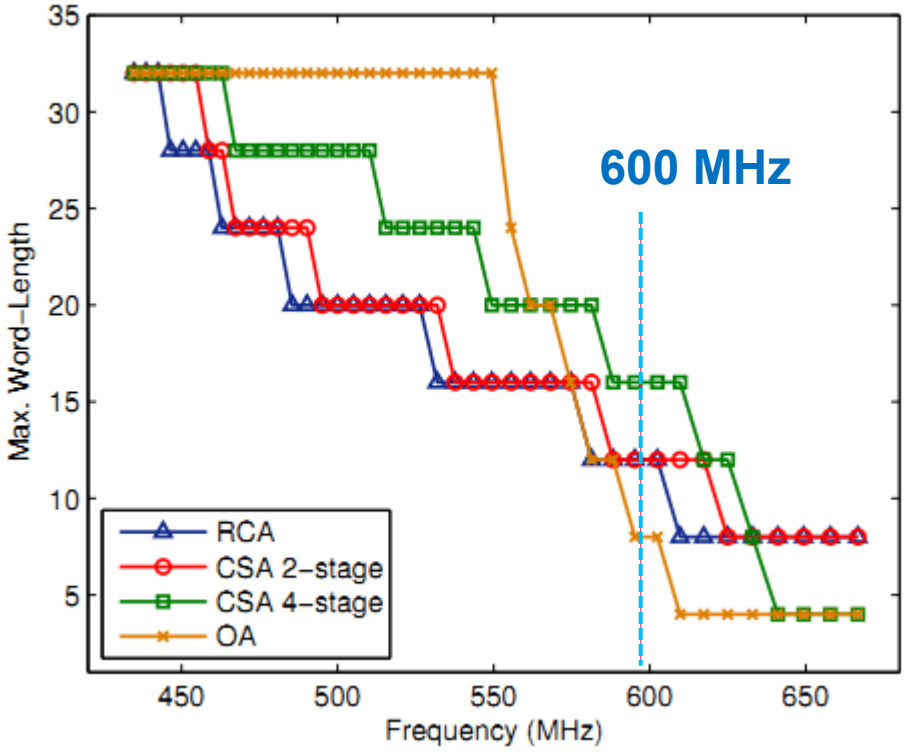
| Freq (MHz) | RCA | CSA-2 | CSA-4 | Online Adder |
|---|---|---|---|---|
| 500 | 20 | 20 | 28 | 32 |
| 550 | 16 | 16 | 20 | 32 |
| | | | | |

CSA and OA are designed to run faster than RCA

Imperial College London

# Design Trade-offs: Accuracy vs Frequency

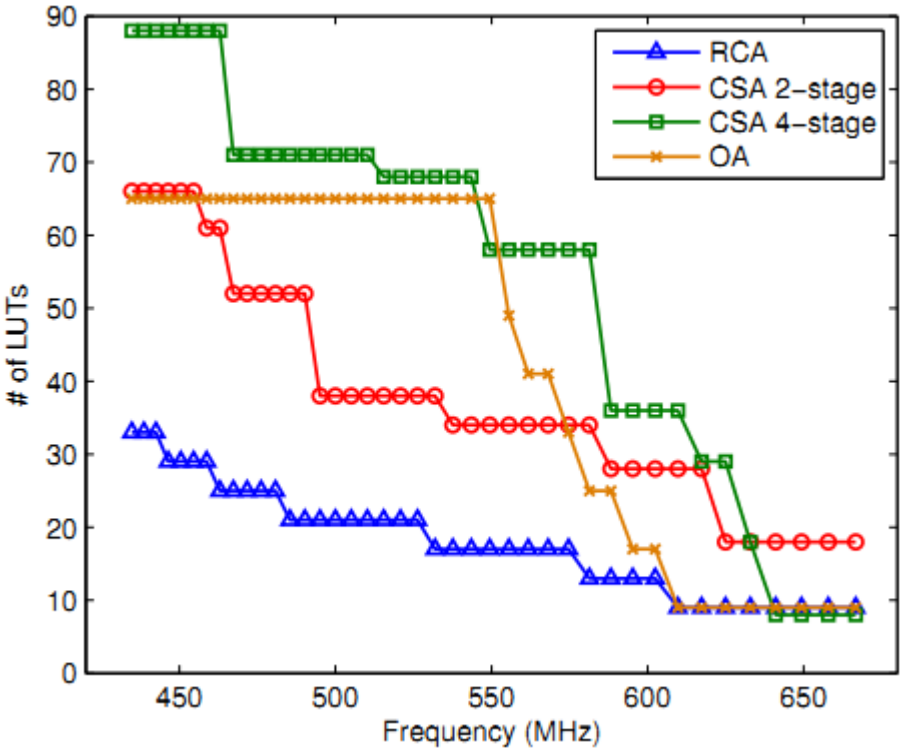
- Original word-length = 32 bits




| Max. Word-length | | | | |
|---|---|---|---|---|
| Freq (MHz) | RCA | CSA-2 | CSA-4 | Online Adder |
| 500 | 20 | 20 | 28 | 32 |
| 550 | 16 | 16 | 20 | 32 |
| 600 | 12 | 12 | 16 | 8 |

# Design Trade-offs: Area vs Frequency

## Area vs Frequency




## Area Overhead than RCA

| Freq (MHz) | CSA-2 | CSA-4 | Online Adder |
|---|---|---|---|
| 500 | 1.8x | 3.4x | 3.1x |
| 550 | 2.0x | 3.4x | 3.8x |
| 600 | 2.2x | 2.8x | 1.3x |

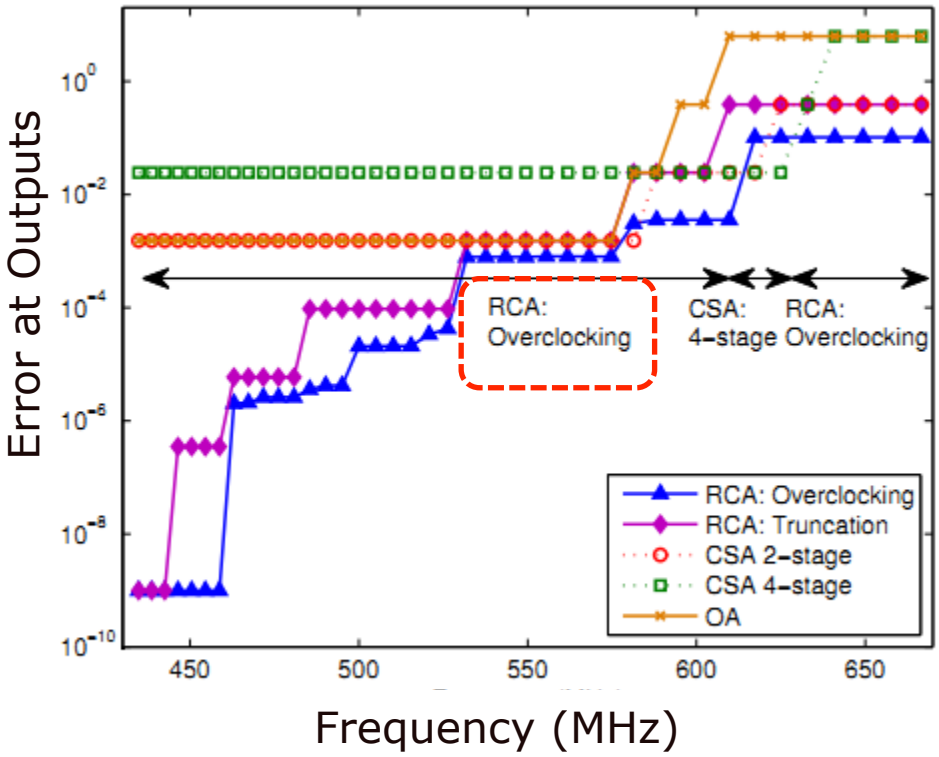
# Outline

- Motivation
- Background
- Design trade-offs for adders
- **Evaluation of the optimum adder design choice**
  - **Example: limited area budget**
  - **Results with a variety of area constraints**
- Conclusion

Imperial College London

# Design Trade-offs: Limited Area Budget

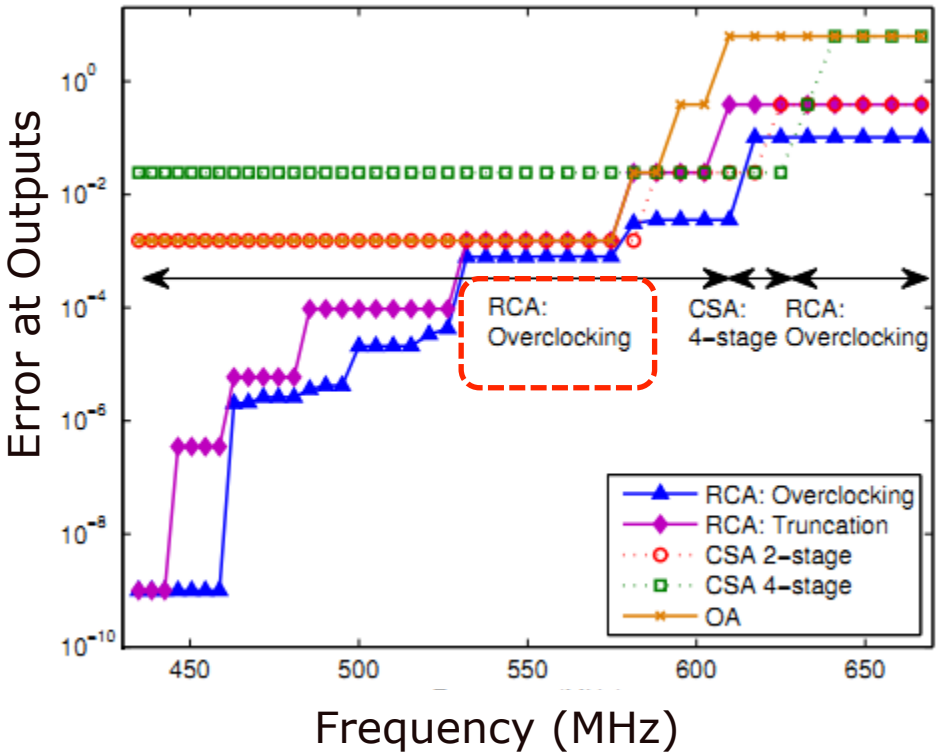
- Optimal design achieves **minimum error**
- Available LUT = 35: RCA with overclocking

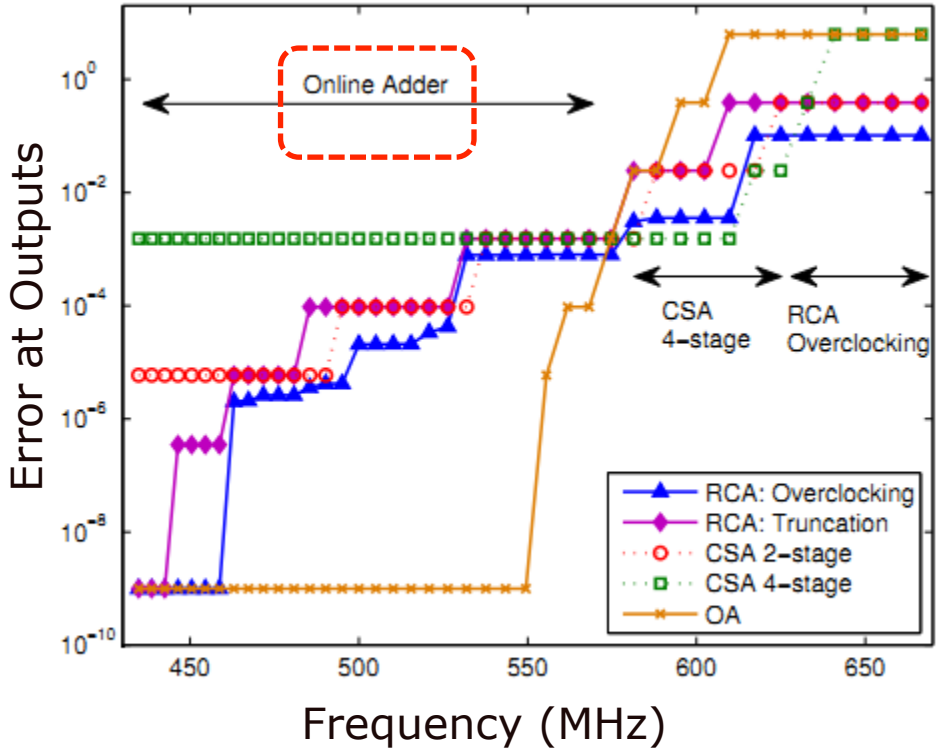


Imperial College London

# Design Trade-offs: Limited Area Budget

- Optimal design achieves **minimum error**
- Available LUT = 35: RCA with overclocking
- Available LUT = 55: Online adder

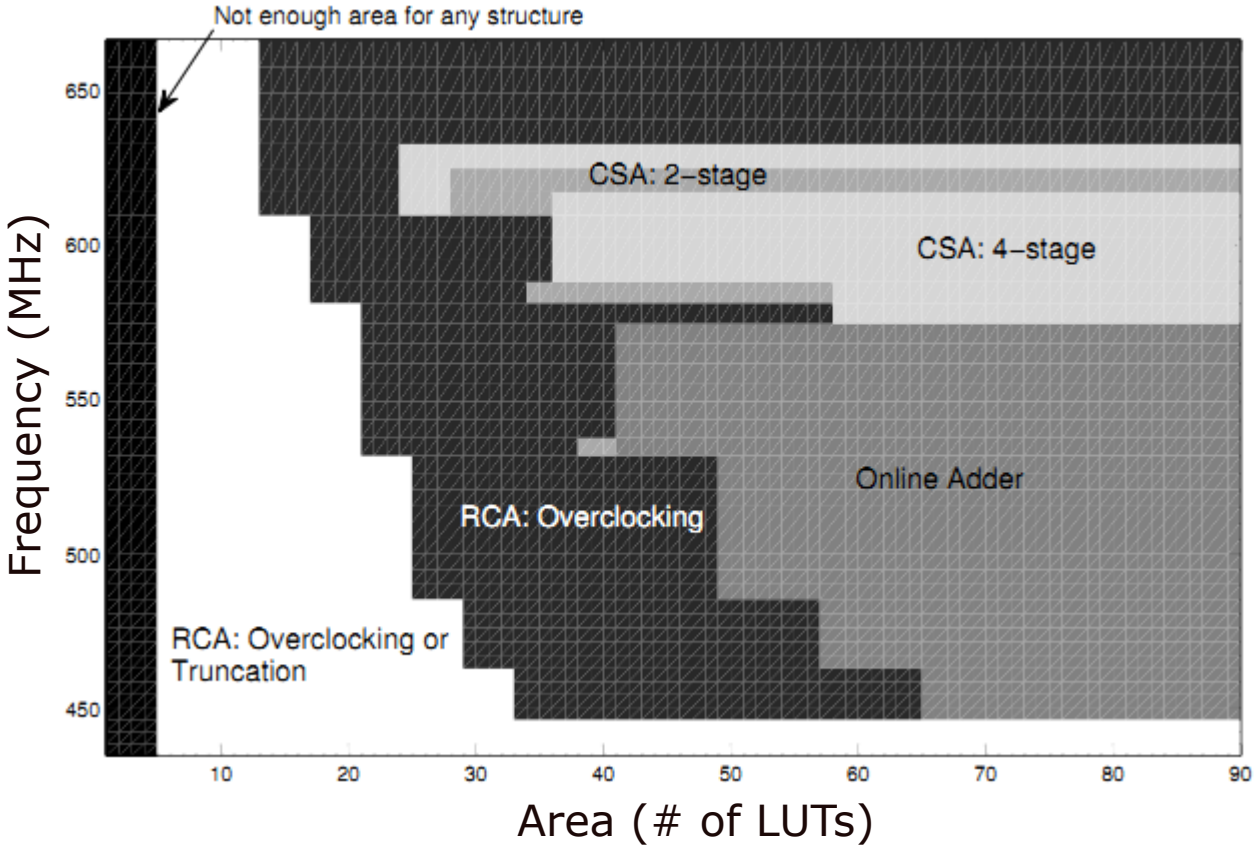


Imperial College London

# Optimum Adder Design Choice

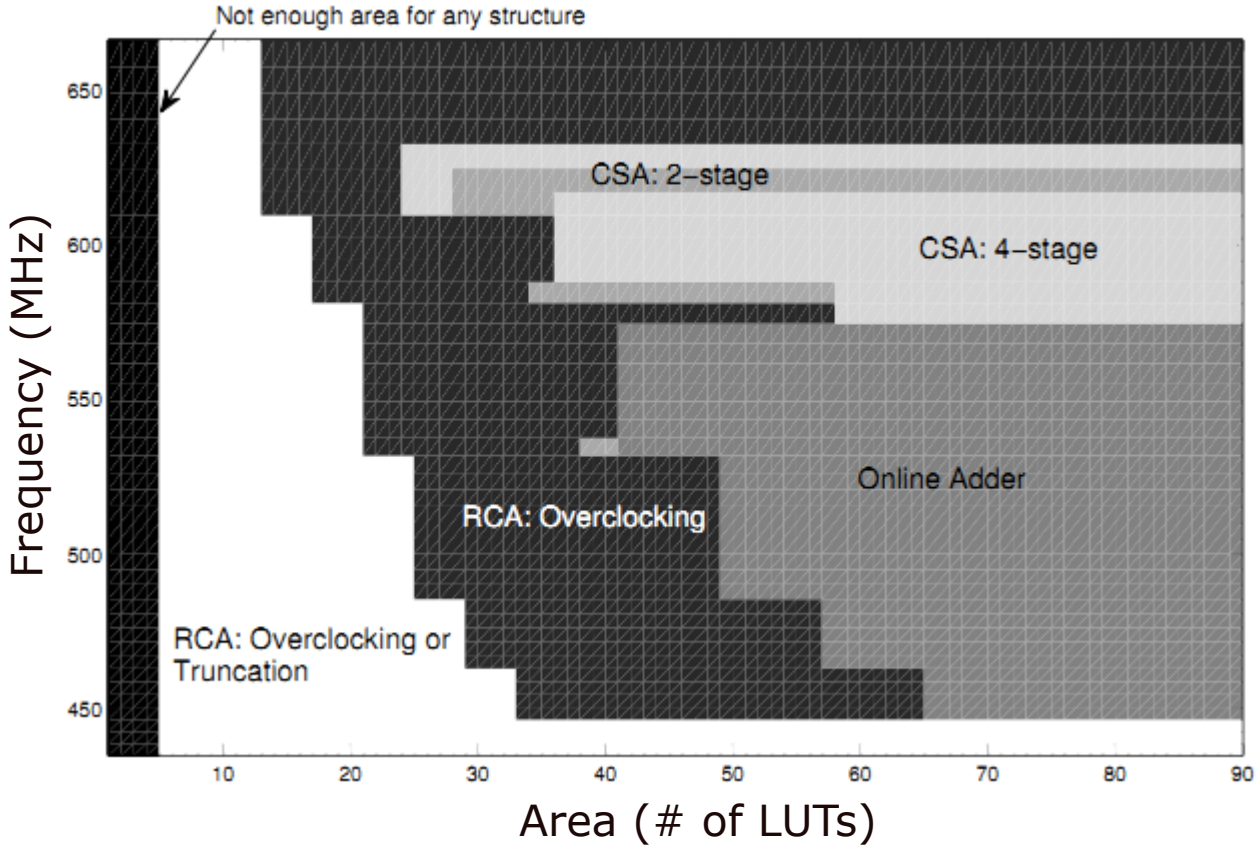
- Optimal design achieves **minimum error**



Imperial College London

# Optimum Adder Design Choice

- Optimal design achieves **minimum error**
- Optimal design achieves **fastest frequency**: see paper



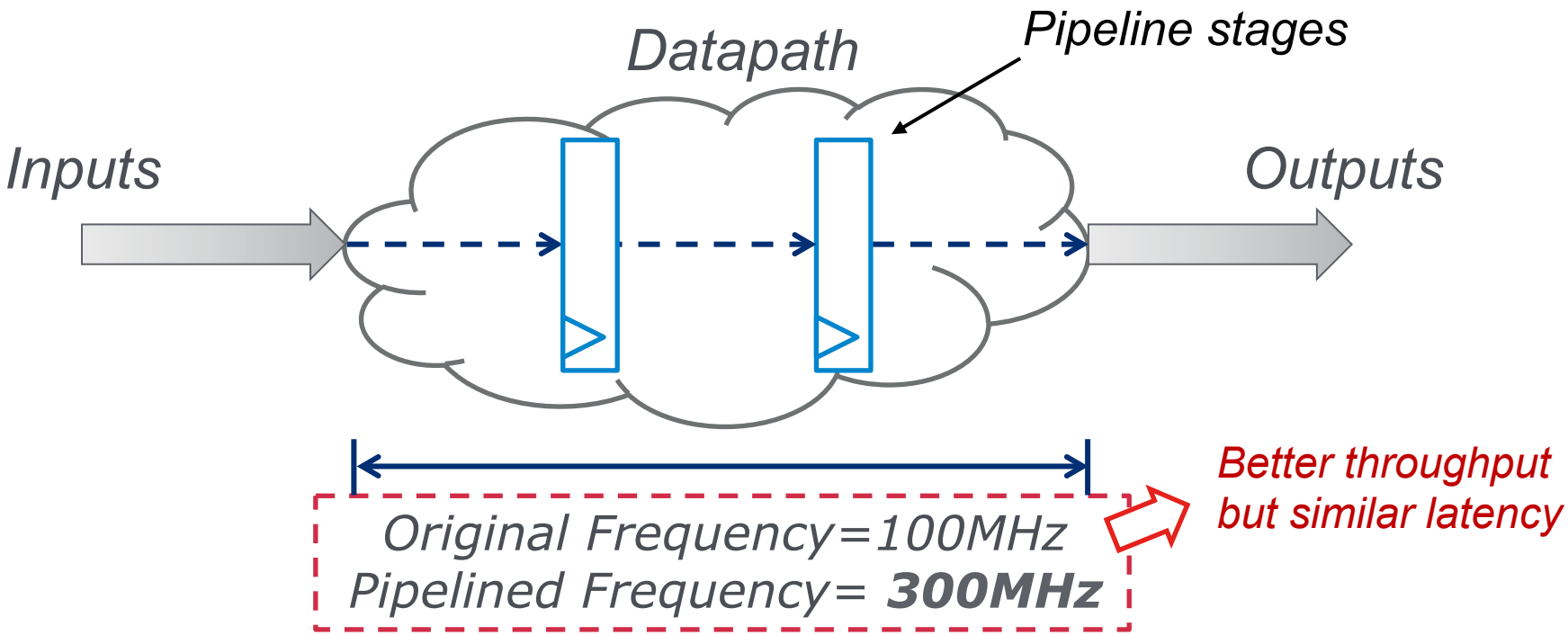
# Conclusion

- We evaluated 3 adder structures: ripple carry adder, carry-select adder and online adder
- We considered 2 design scenarios that trade accuracy for performance: overclocking and truncation
- We show that
  - Limited area budget: using RCA with overclocking
  - Relaxed area budget: using online adder
- Details in papers and poster

Imperial College
London

# It is all about Performance

- FPGAs run at much lower frequency than CPUs and GPUs
  - Pipelining? : not helpful in reducing datapath latency



# It is all about Performance--backup

- FPGAs run at much lower frequency than CPUs and GPUs
  - Pipelining? : not helpful in reducing datapath latency
- EDA tools provide conservative timing to cover variations