Avanergy: An Architecture-VLSI Abstraction for Automated Energy-Aware Design-Space Exploration Tools

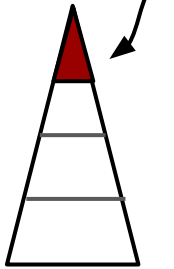
Christopher Torng

University of Southern California

A very different world of hardware design may be emerging

Comparing two different profiles of hardware designers

1

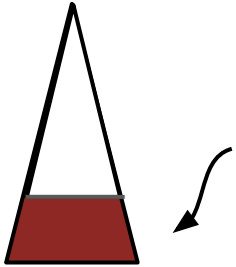


Hardware experts with ~5-20+ years of experience

CPU

TPU

Accelerators



... who are also

domain enthusiasts with ~1 year or less of experience

A specific way that things have always been done

Specification

RTL Design

Verification

Physical Design and Impl

Development Time

Tapeout

Software Stack

Architecture

VLSI

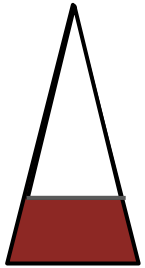
Technology

A specific design methodology

A very different world of hardware design may be emerging

Comparing two different profiles of hardware designers

2

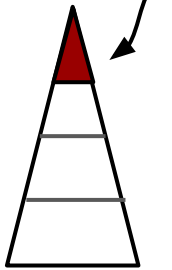


Hardware designers

with ~1 year or
less of experience

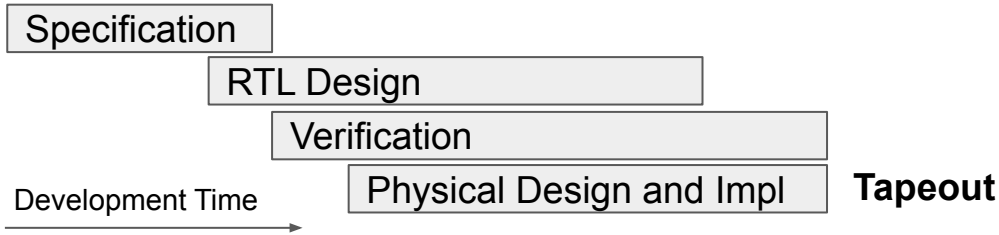
... whose primary jobs are

Machine Learning
Image Processing
Video Coding
Cryptography
Wireless

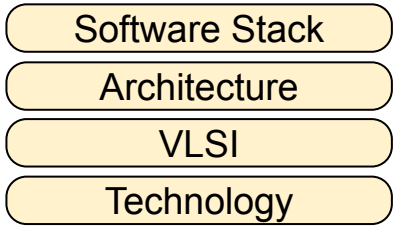


Domain Experts

Key Question: Do our existing methods work?



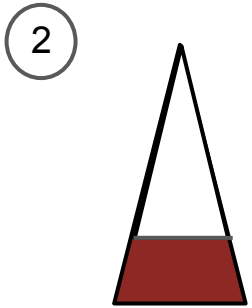
Tapeout



Does the stack need to change to "hide" more?

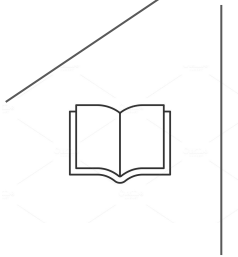
A very different world of hardware design may be emerging

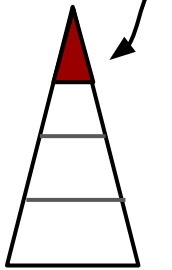
Comparing two different profiles of hardware designers



Hardware designers

with ~1 year or
less of experience





... whose primary jobs are

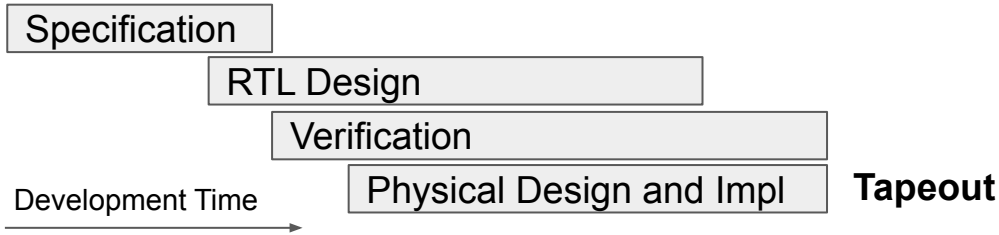
Machine Learning
Image Processing
Video Coding
Cryptography
Wireless

Domain Experts

ChatGPT

Chip-Chat (2023)
arXiv:2305.13243

Key Question: Do our existing methods work?

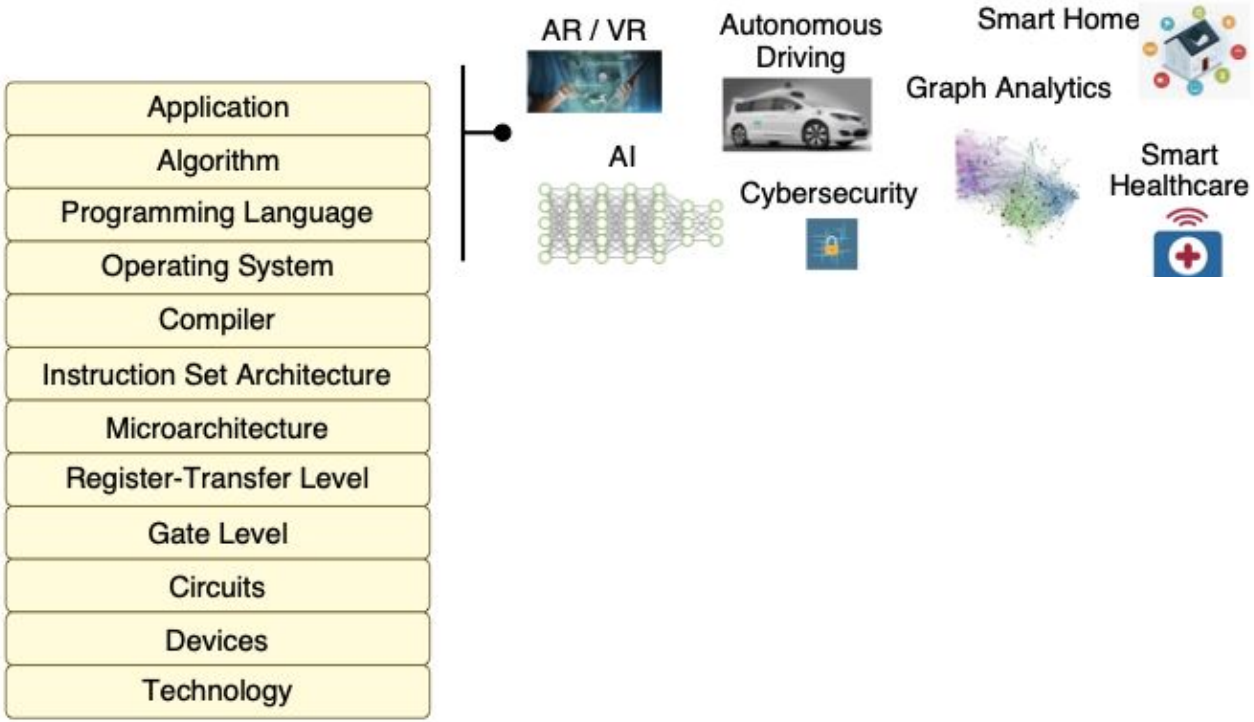


Tapeout

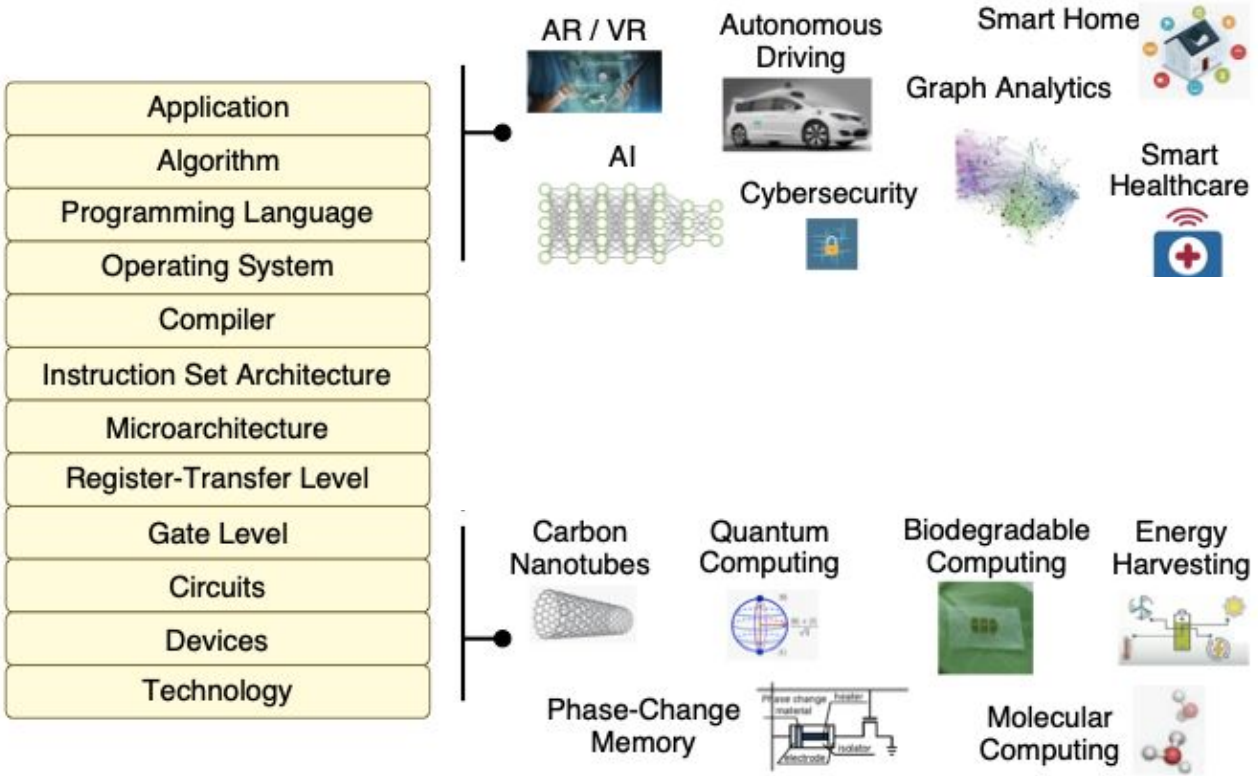
Software Stack
Architecture
VLSI
Technology

Does the stack need to change to "hide" more?

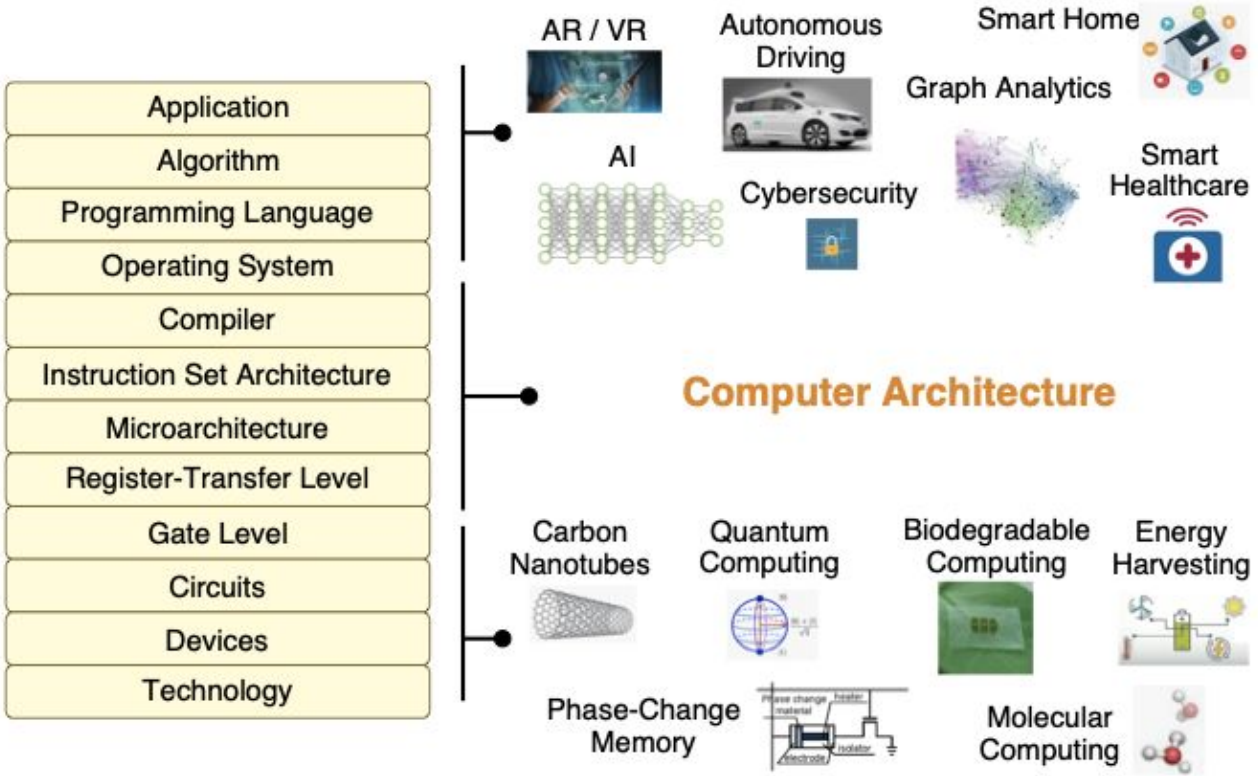
If properly empowered, this new audience could transform the hardware design community



If properly empowered, this new audience could transform the hardware design community



If properly empowered, this new audience could transform the hardware design community



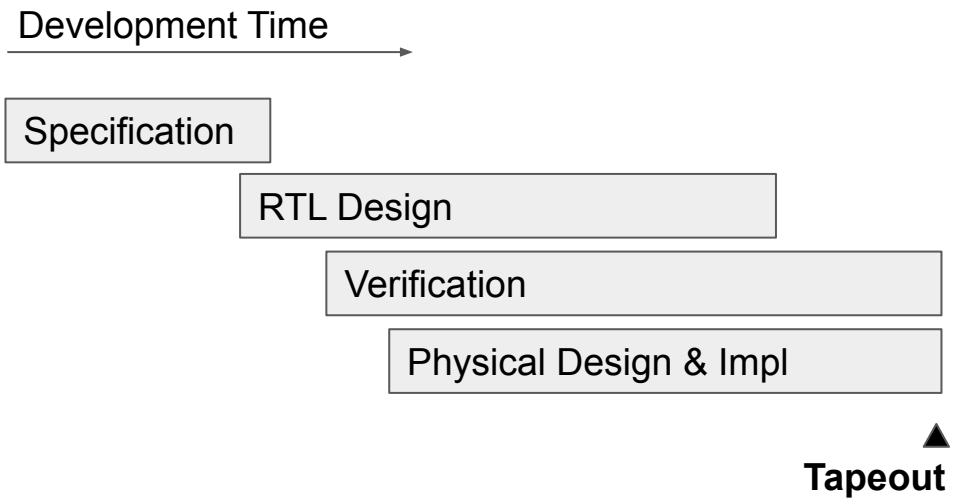
Quick Summary of New Audience

We may have **more people** ..

.. with **less hardware expertise** ..

.. who are relying heavily on **automated tools**

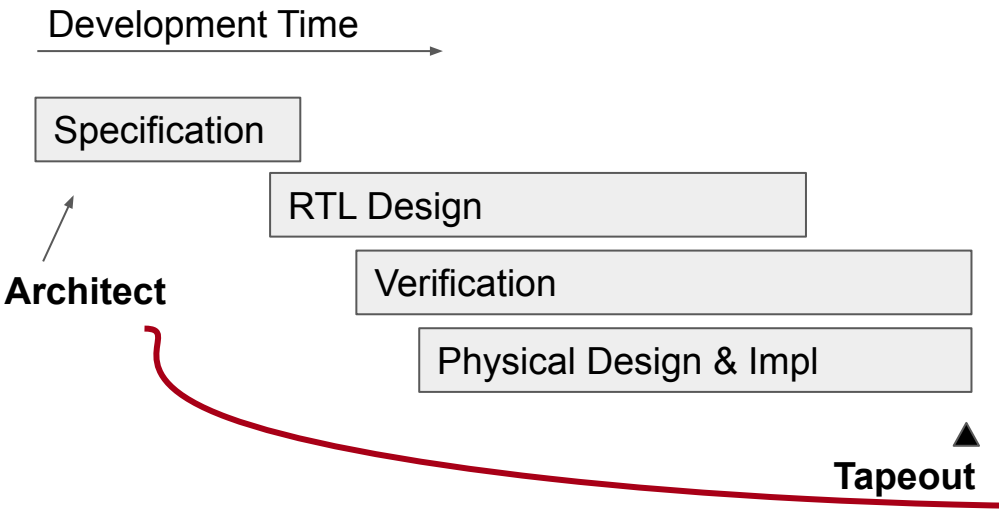
Insight #1: Traditional design methodologies are built for our current profiles of expertise



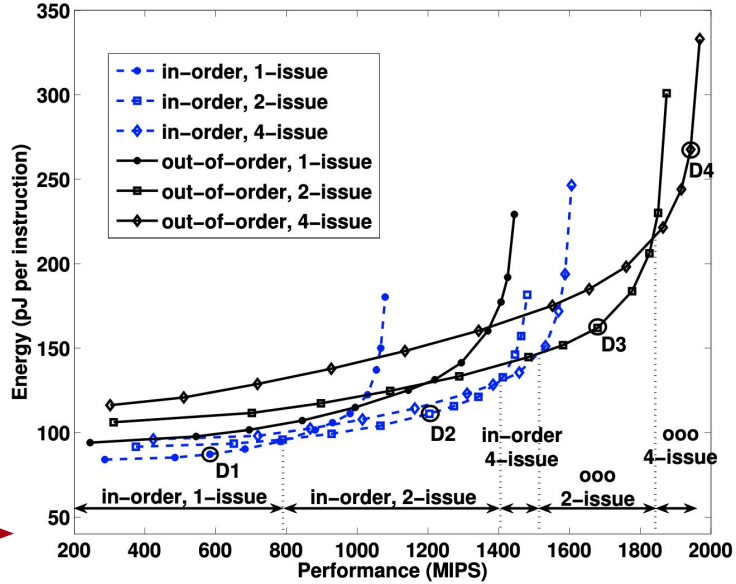
Traditional Chip Development

- An architect comes up with an architectural design specification
- The RTL design team takes the specification and builds the design
- In parallel, the VLSI design team moves through physical impl and iterates with RTL for refinement
- Tapeout

Insight #1: Traditional design methodologies are built for our current profiles of expertise

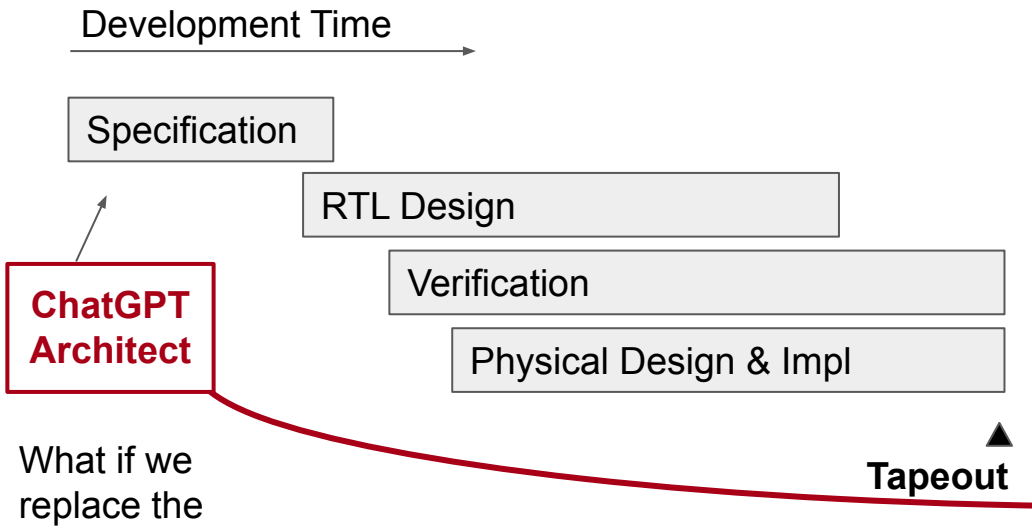


Problem - The architect is assumed to be a tremendous expert, and can pick <u>one good design point</u> among many ...



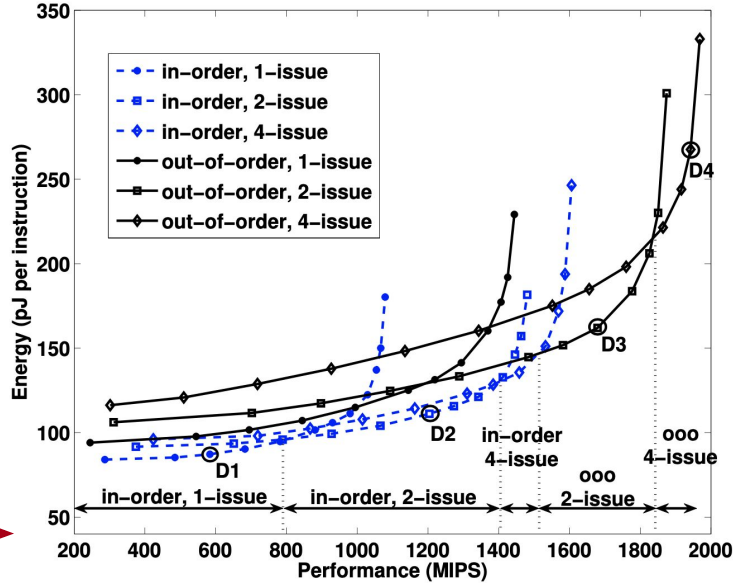
Example of Design Space
Azizi et al. (ISCA 2010)

Insight #1: Traditional design methodologies are built for our current profiles of expertise



What if we replace the architect?

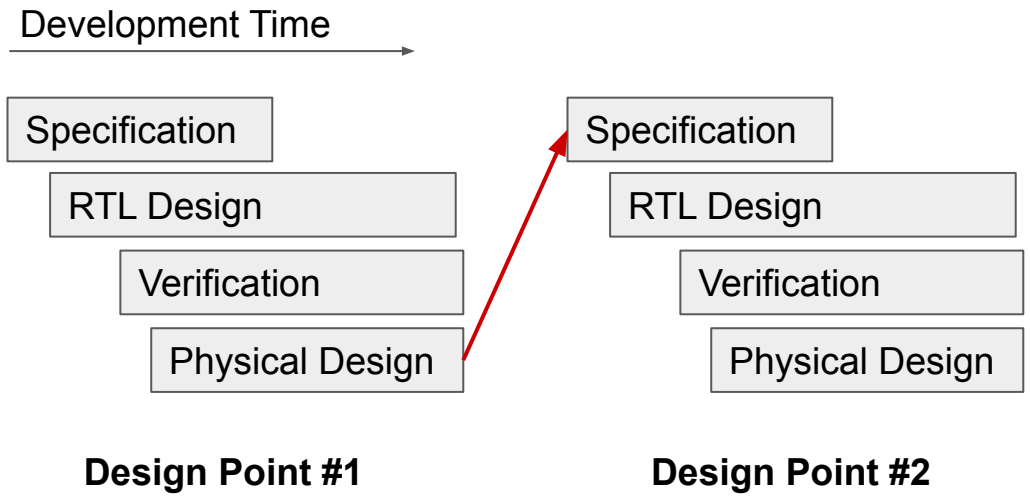
Problem - The architect is assumed to be a tremendous expert, and can pick <u>one good design point</u> among many ...



Example of Design Space
Azizi et al. (ISCA 2010)

Insight #2: Iterative approaches can reduce the burden of getting things right the first time

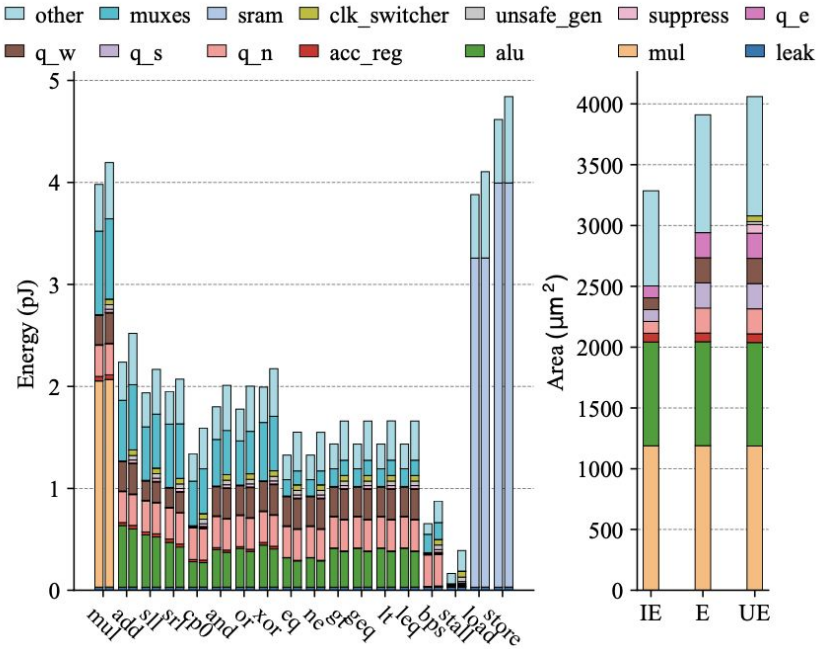
- Multiple iterations leaves room to get it wrong and then correct the mistakes
- But the iterations must then become **very short** in order to fit multiple of them
- This exactly describes an agile hardware design methodology!



Insight #3: Automation requires having systematic approaches built around quantifiable abstractions

Example question:

- How good of an energy efficiency result is good enough?



Example of Energy Breakdown Plot in 28nm Technology
Torng et al. (HPCA 2021)

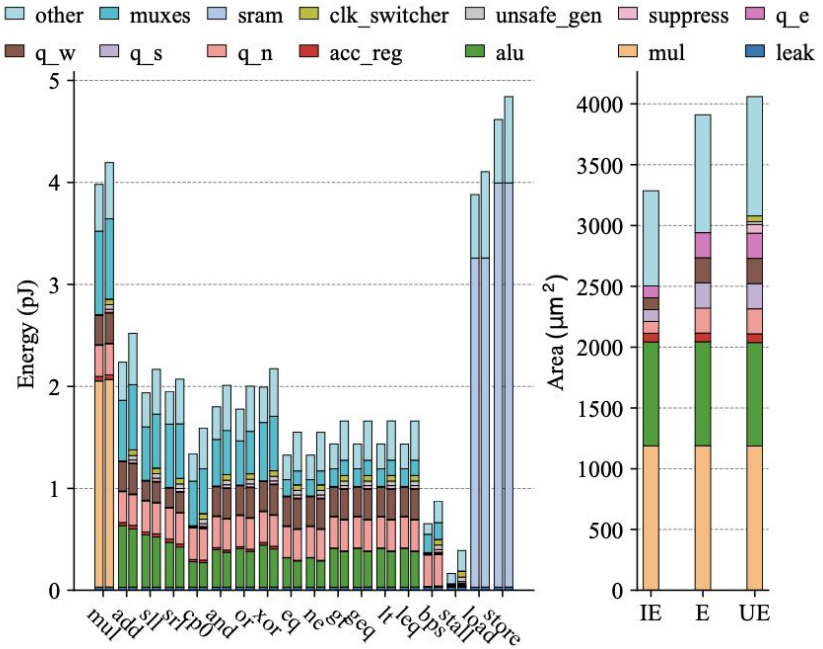
Insight #3: Automation requires having systematic approaches built around quantifiable abstractions

Example question:

- How good of an energy efficiency result is good enough?

In an fully automated flow, there is no place for allowing designs to pass review that "look okay"

There must be a **quantifiable constraint**, as well as a systematic approach for generating these constraints and enforcing them.



Example of Energy Breakdown Plot in 28nm Technology
Torng et al. (HPCA 2021)

Insight #3: Automation requires having systematic approaches built around quantifiable abstractions

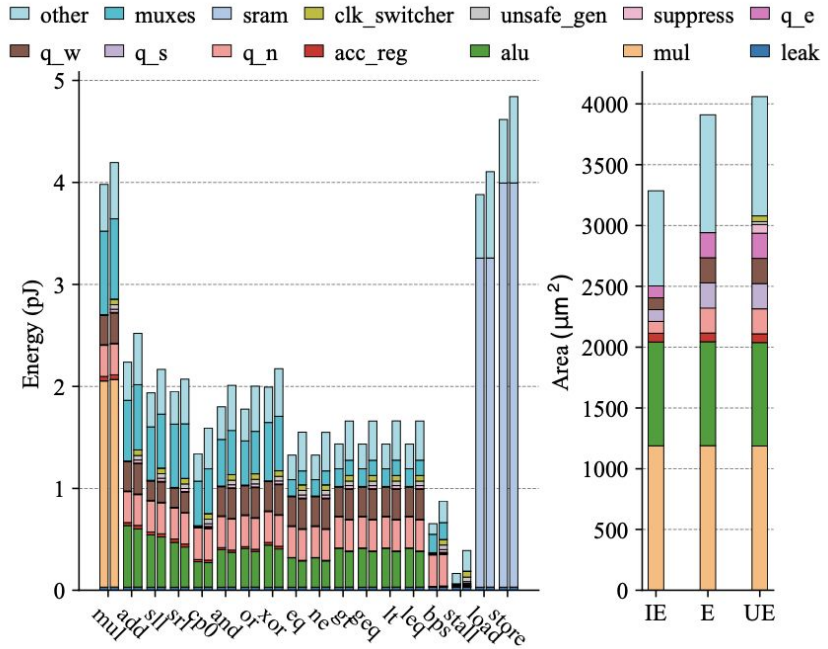
Example question:

- How good of an energy efficiency result is good enough?

In an fully automated flow, there is no place for allowing designs to pass review that "look okay"

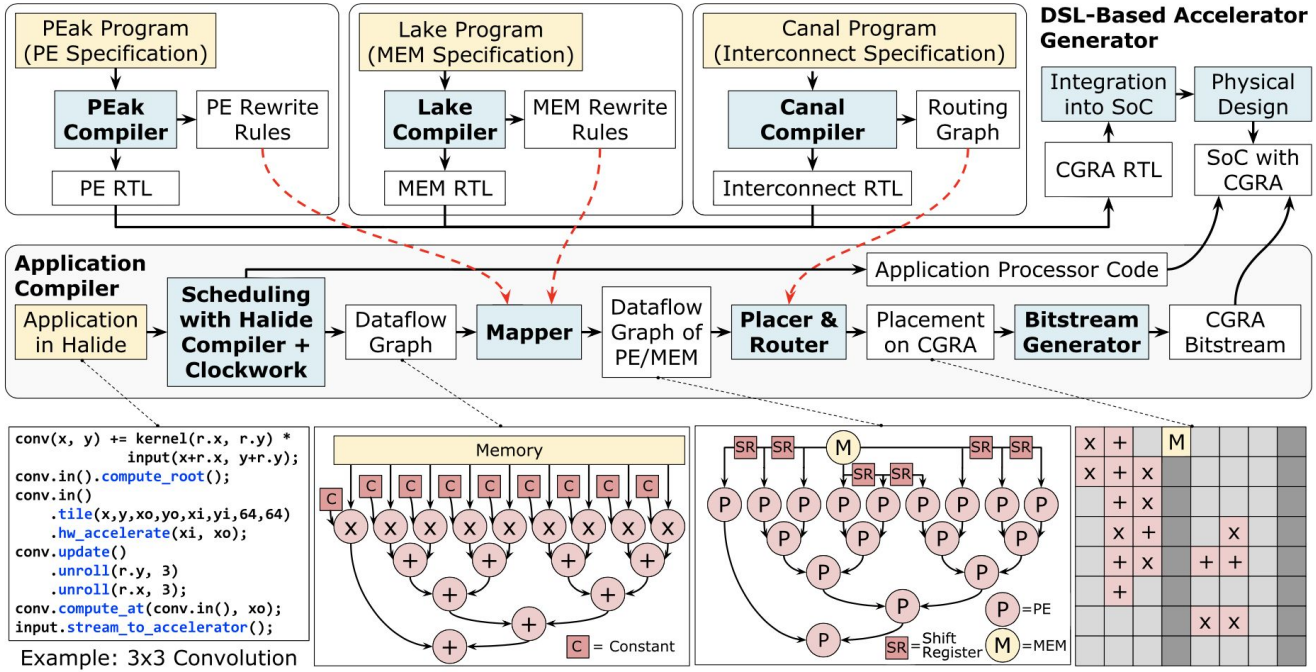
There must be a **quantifiable constraint**, as well as a systematic approach for generating these constraints and enforcing them.

Expert – Who will you hire to tell you it looks okay?



Example of Energy Breakdown Plot in 28nm Technology
Torng et al. (HPCA 2021)

Insight #4: Some aspects of complex SoC design pose a far more critical challenge than others

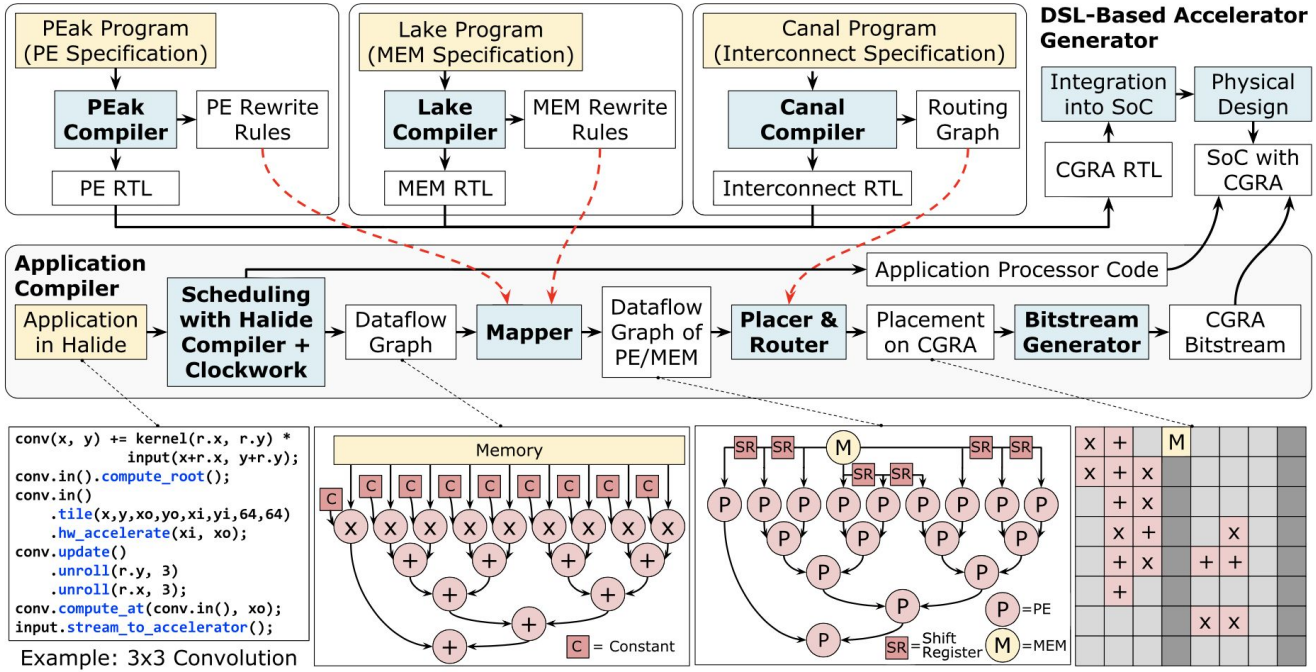


Case for Focusing on Energy

Energy modeling is becoming **more challenging** at exactly the same time that automated approaches for energy are becoming **increasingly valuable**

- Scheduler – Different kernel scheduling
- Compiler – Different mappings
- Architecture – Different compute and memory organization
- VLSI - Ground truth

Insight #4: Some aspects of complex SoC design pose a far more critical challenge than others



Case for Focusing on Energy

It is such a challenge that even experts do not want to do it

- TECS 2023
- TODAES 2022
- MICRO 2022
- IEEE MICRO 2020
- ICCAD 2019

Example - A series of very efficient academic CGRAs are marketed as"energy minimal", but they are in fact just "more efficient" than others

- MICRO 2022
- ISCA 2021

Our system-level requirements help shape our thinking

- How much should we refine an RTL design's energy efficiency to be "good enough"?

Note #1: Can you try to answer these questions as a hardware expert?

Our system-level requirements help shape our thinking

- How much should we refine an RTL design's energy efficiency to be "good enough"?

- Isn't lower energy always better?

Note #1: Can you try to answer these questions as a hardware expert?

Our system-level requirements help shape our thinking

- How much should we refine an RTL design's energy efficiency to be "good enough"?

- Isn't lower energy always better?

- What is a meaningful definition of an **energy constraint**?

Note #1: Can you try to answer these questions as a hardware expert?

Note #2: Notice that we also begin to introduce new terminology ... energy constraints

Our system-level requirements help shape our thinking

- How much should we refine an RTL design's energy efficiency to be "good enough"?

- Isn't lower energy always better?

- What is a meaningful definition of an **energy constraint**?

Note #1: Can you try to answer these questions as a hardware expert?

Note #2: Notice that we also begin to introduce new terminology ... energy constraints

Our system-level requirements help shape our thinking

- How much should we refine an RTL design's energy efficiency to be "good enough"?
- Isn't lower energy always better?
- What is a meaningful definition of an **energy constraint**?
- What does it mean if we have violated energy constraints?
- How can I characterize a design to see if I meet an energy constraint or not?
- What are the most meaningful semantics we can assign to the idea of "meeting energy"?

Note #1: Can you try to answer these questions as a hardware expert?

Note #2: Notice that we also begin to introduce new terminology ... energy constraints

One other thing that follows from automation

An automated approach requires an **explicitly defined energy model** (cost model)

- Previously we did ***not*** need good energy models!
- Simply bank on the chief architect's experience and intuition ...

One other thing that follows from automation

An automated approach requires an **explicitly defined energy model** (cost model)

- Previously we did ***not*** need good energy models!
- Simply bank on the chief architect's experience and intuition ...

This is one of the key hurdles that faces an automated world

- The energy models that underlie future automation are very important

One other thing that follows from automation

An automated approach requires an **explicitly defined energy model** (cost model)

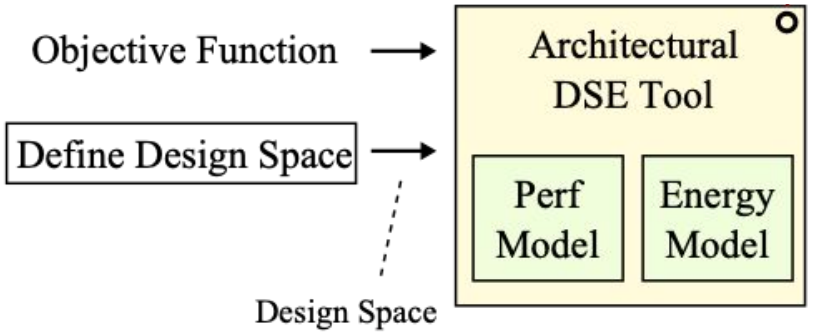
- Previously we did *not* need good energy models!
- Simply bank on the chief architect's experience and intuition ...

This is one of the key hurdles that faces an automated world

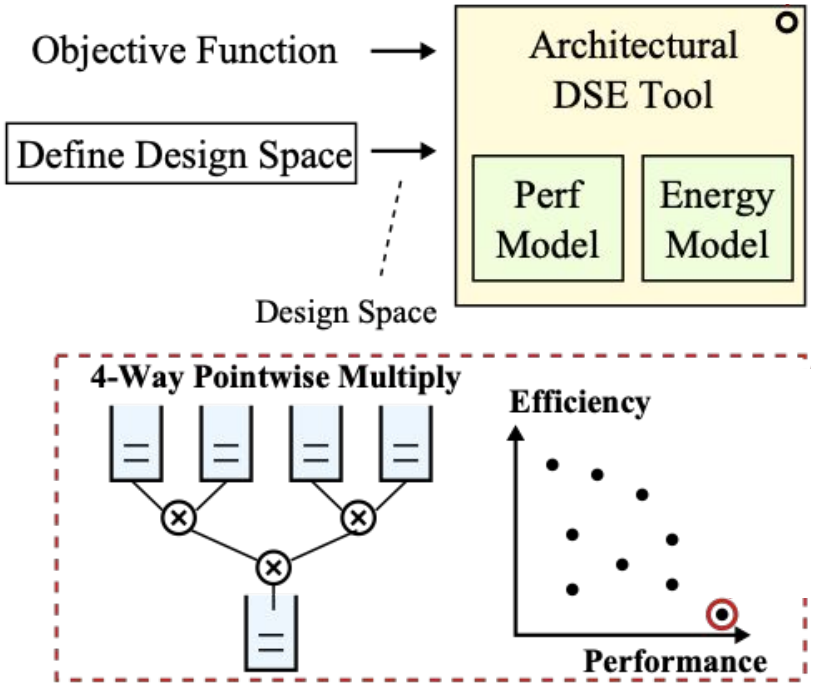
- The energy models that underlie future automation are very important
- But these models can be *buggy*

One of the key goals of this project is to solve this through iteration and a new abstraction

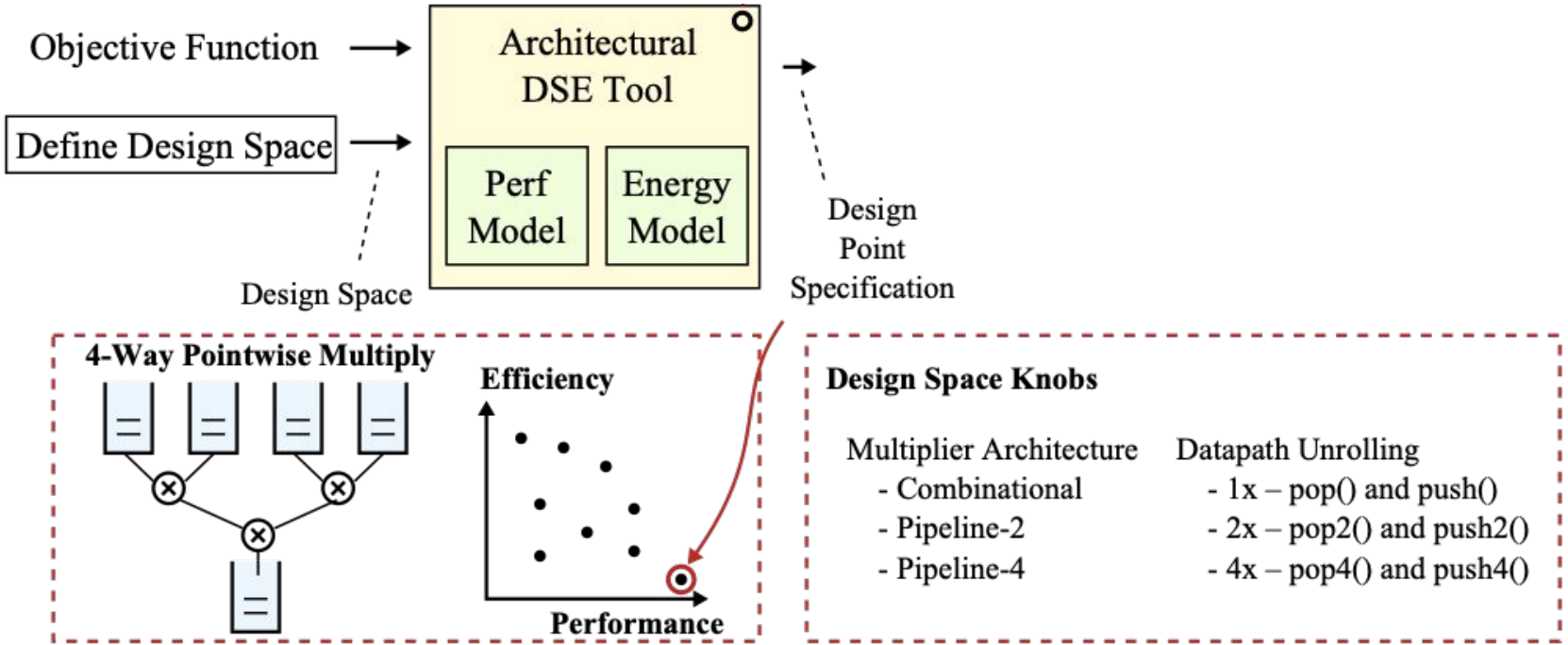
Avanergy: An Architecture-VLSI Abstraction (Overview)



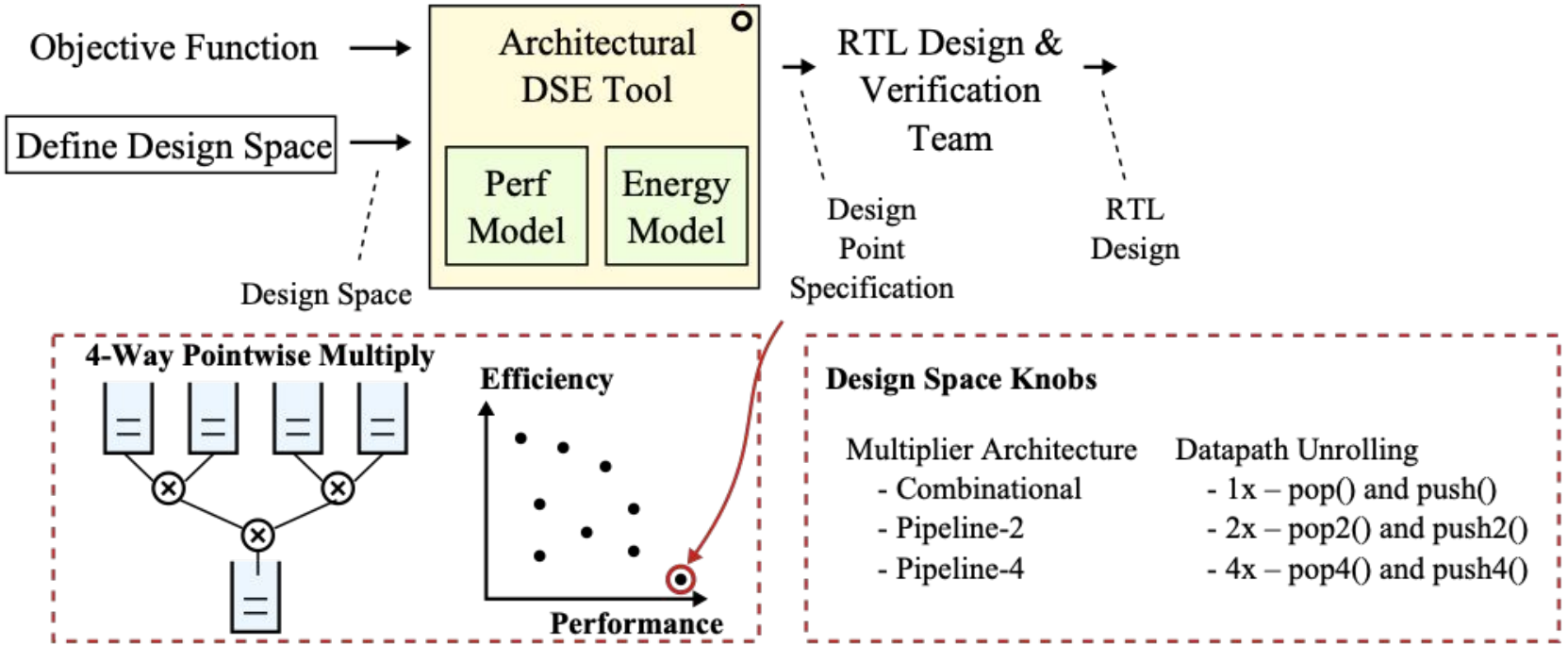
Avanergy: An Architecture-VLSI Abstraction (Overview)



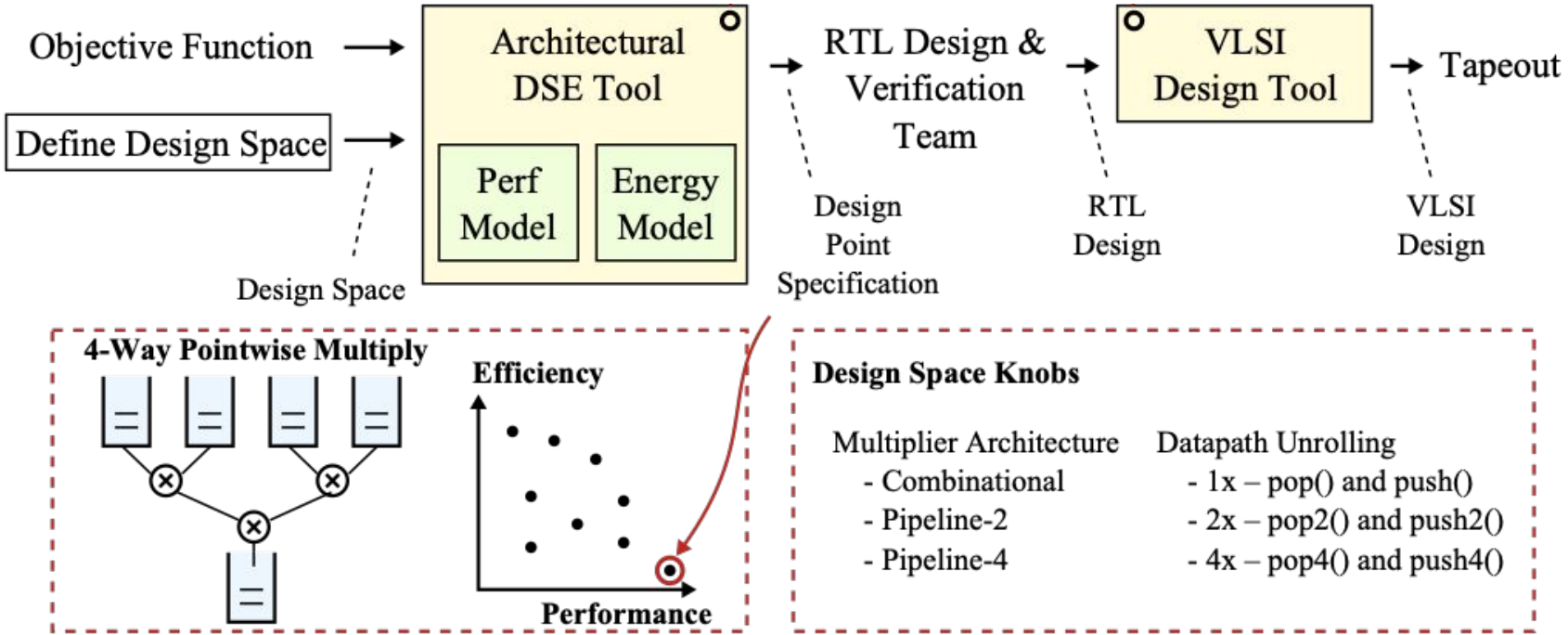
Avanergy: An Architecture-VLSI Abstraction (Overview)



Avanergy: An Architecture-VLSI Abstraction (Overview)

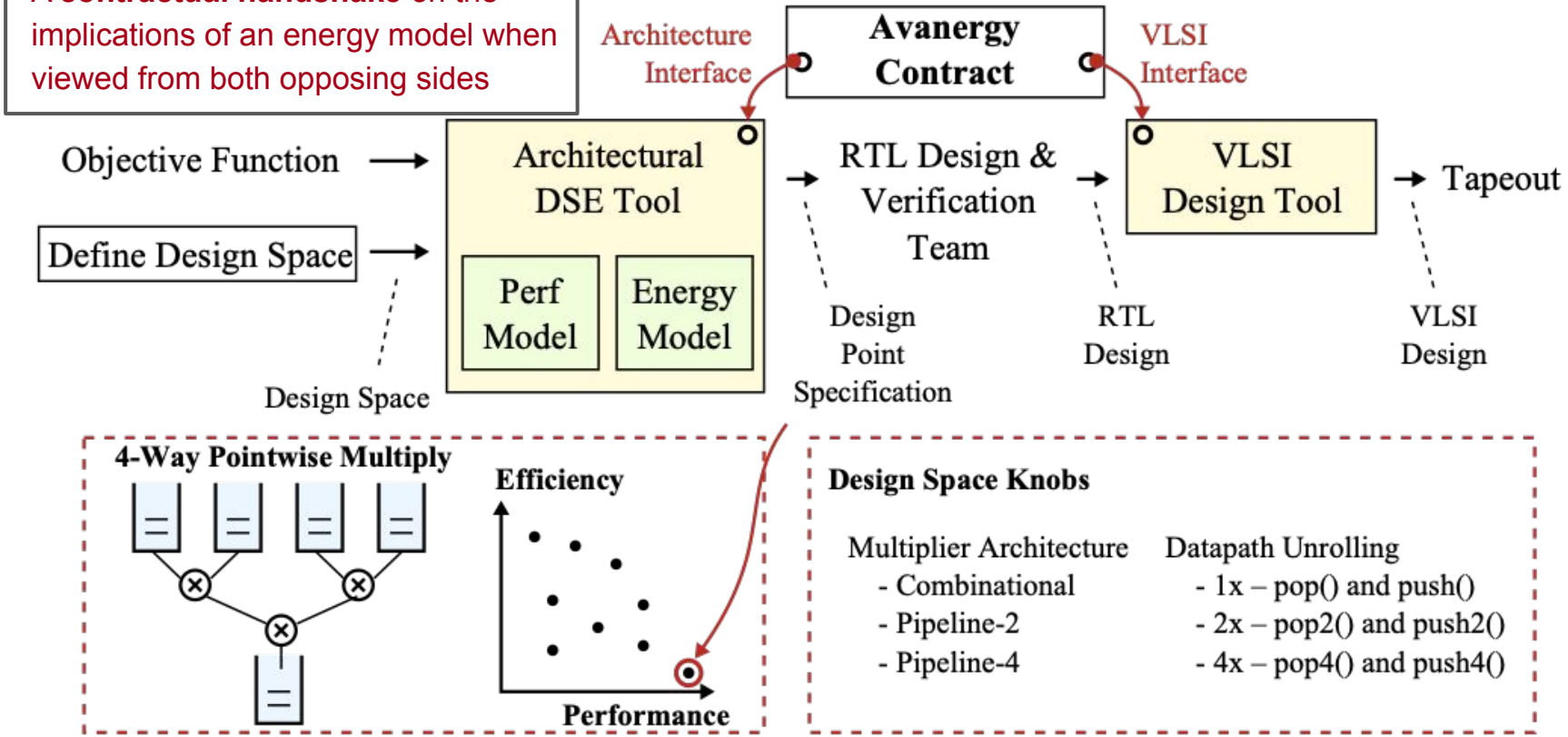


Avanergy: An Architecture-VLSI Abstraction (Overview)

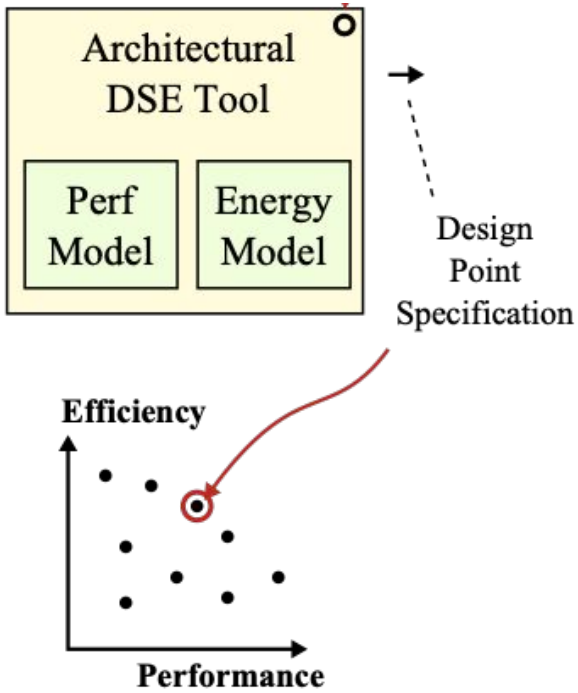


Avanergy: An Architecture-VLSI Abstraction (Overview)

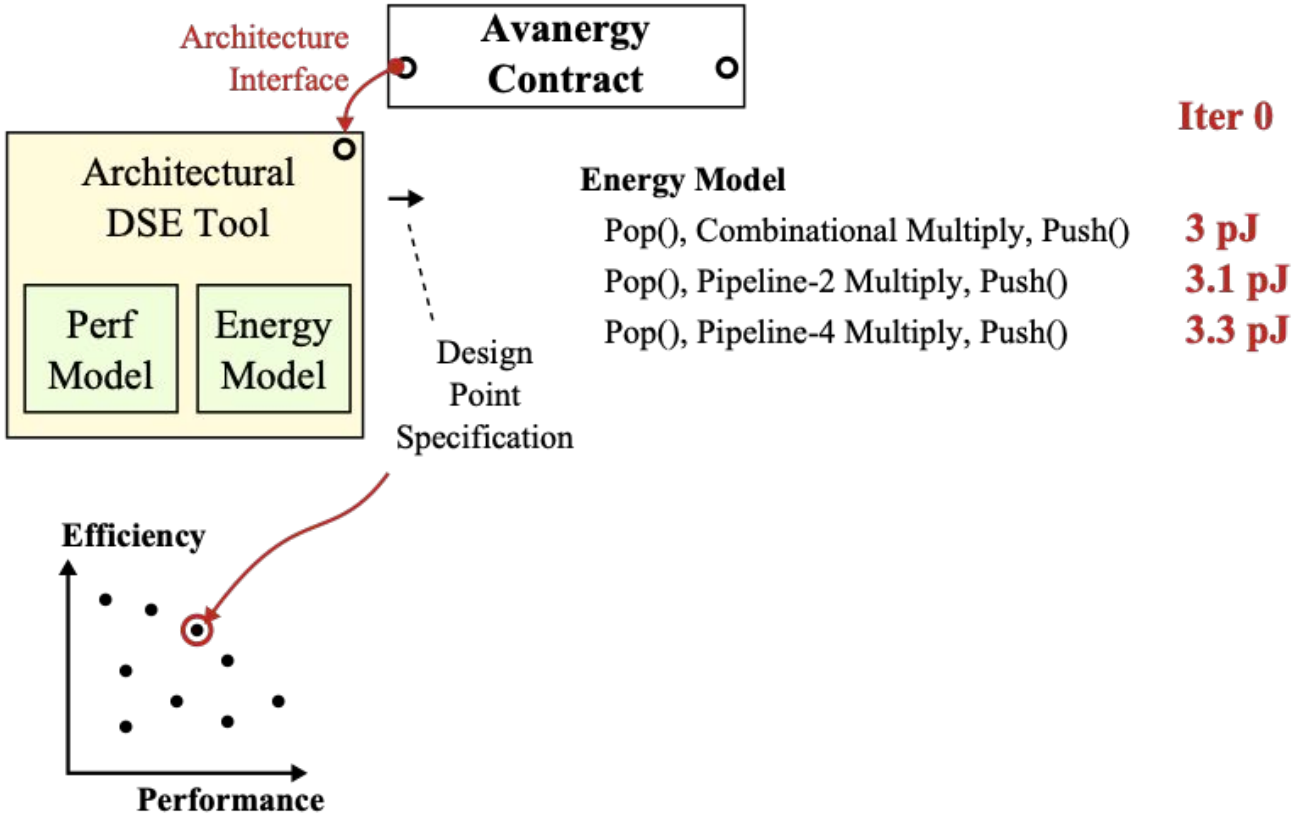
A **contractual handshake** on the implications of an energy model when viewed from both opposing sides



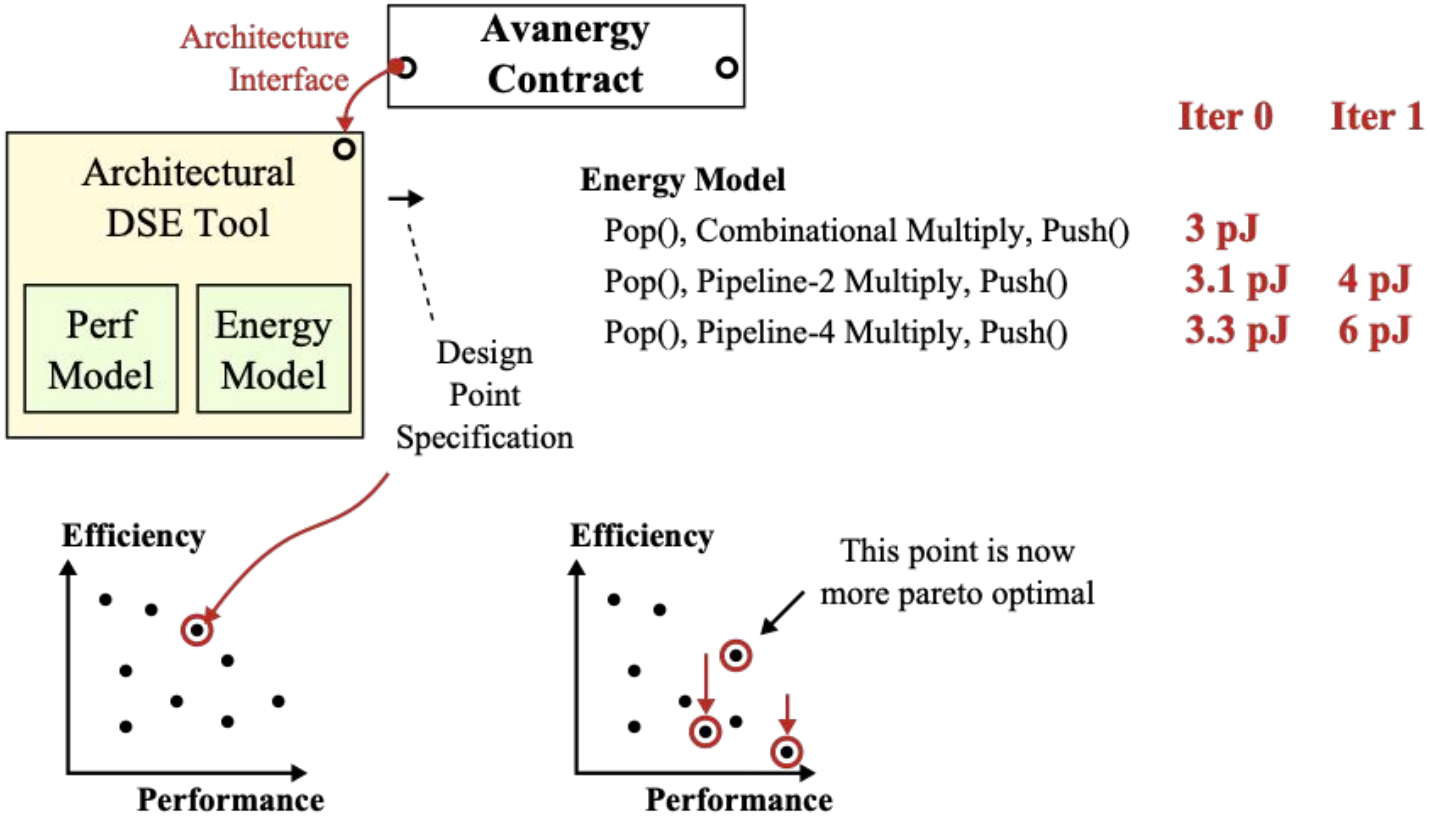
The Avanergy Architecture Interface



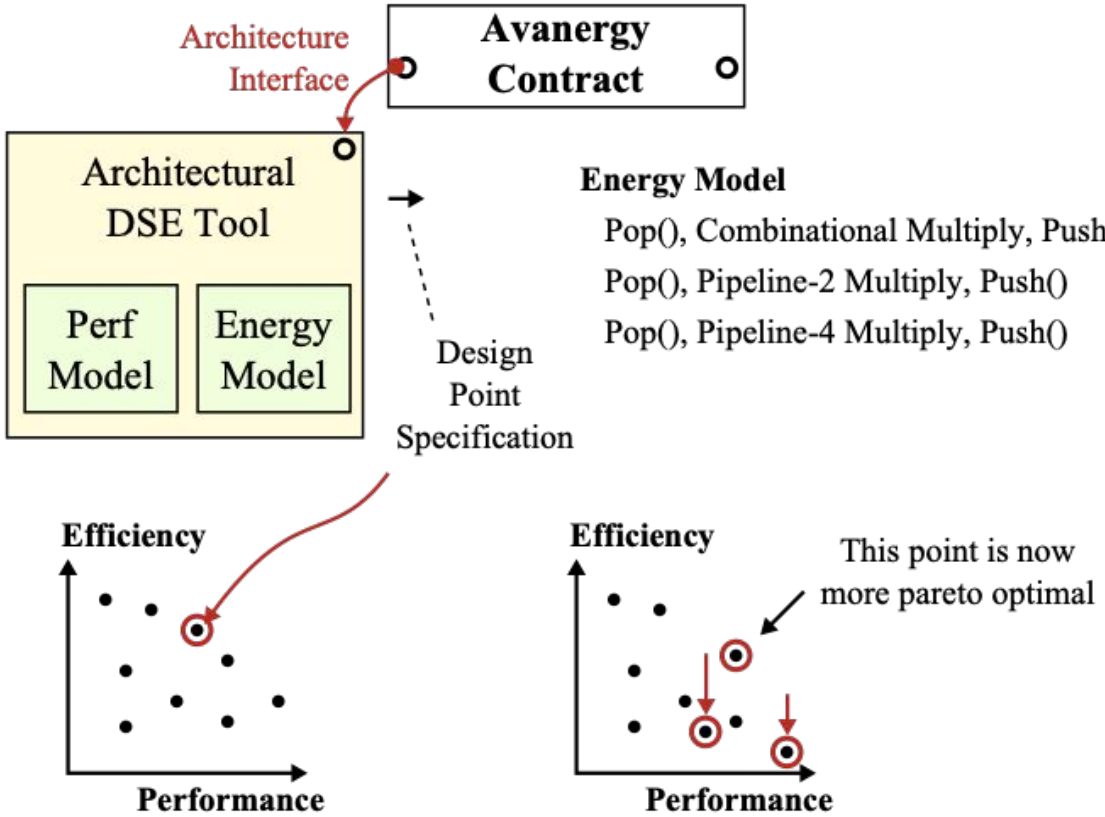
The Avanergy Architecture Interface



The Avanergy Architecture Interface



The Avanergy Architecture Interface

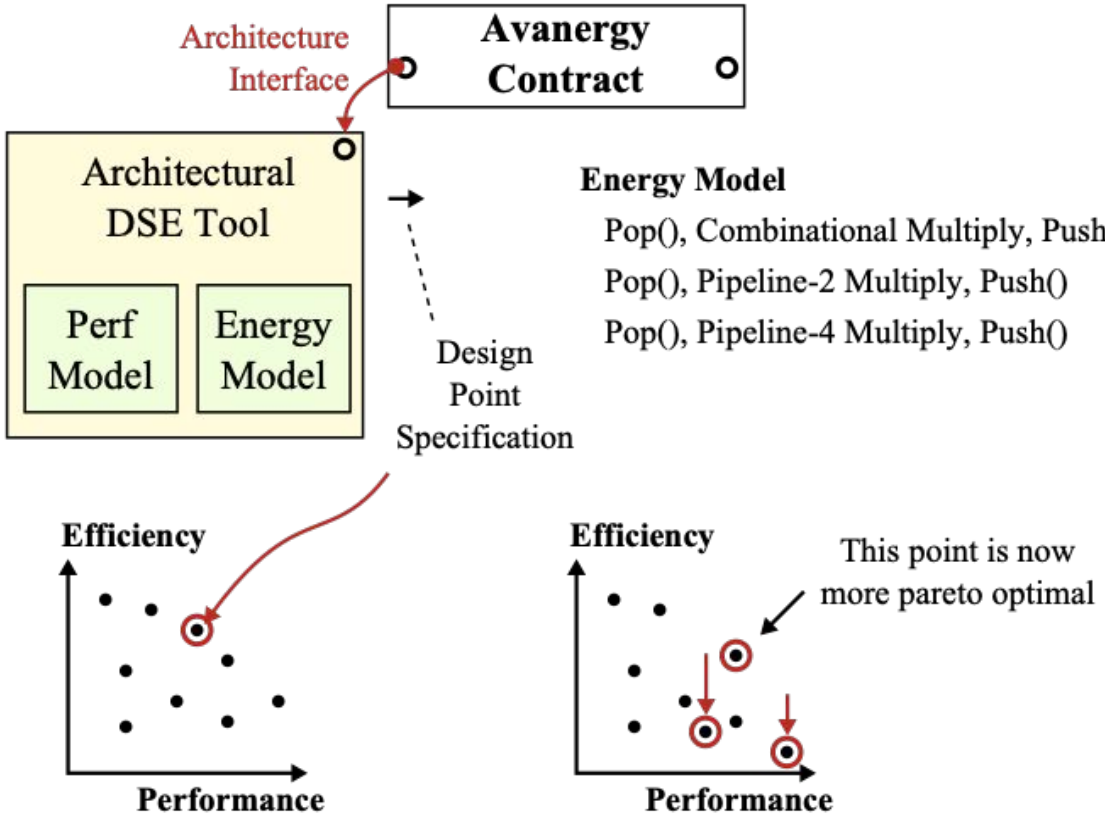


Energy Model	Iter 0	Iter 1
Pop(), Combinational Multiply, Push()	3 pJ	
Pop(), Pipeline-2 Multiply, Push()	3.1 pJ	4 pJ
Pop(), Pipeline-4 Multiply, Push()	3.3 pJ	6 pJ

Goal of the architecture interface

Determine "how wrong" the transaction energy cost must be, where the DSE tool would have instead picked a different design point.

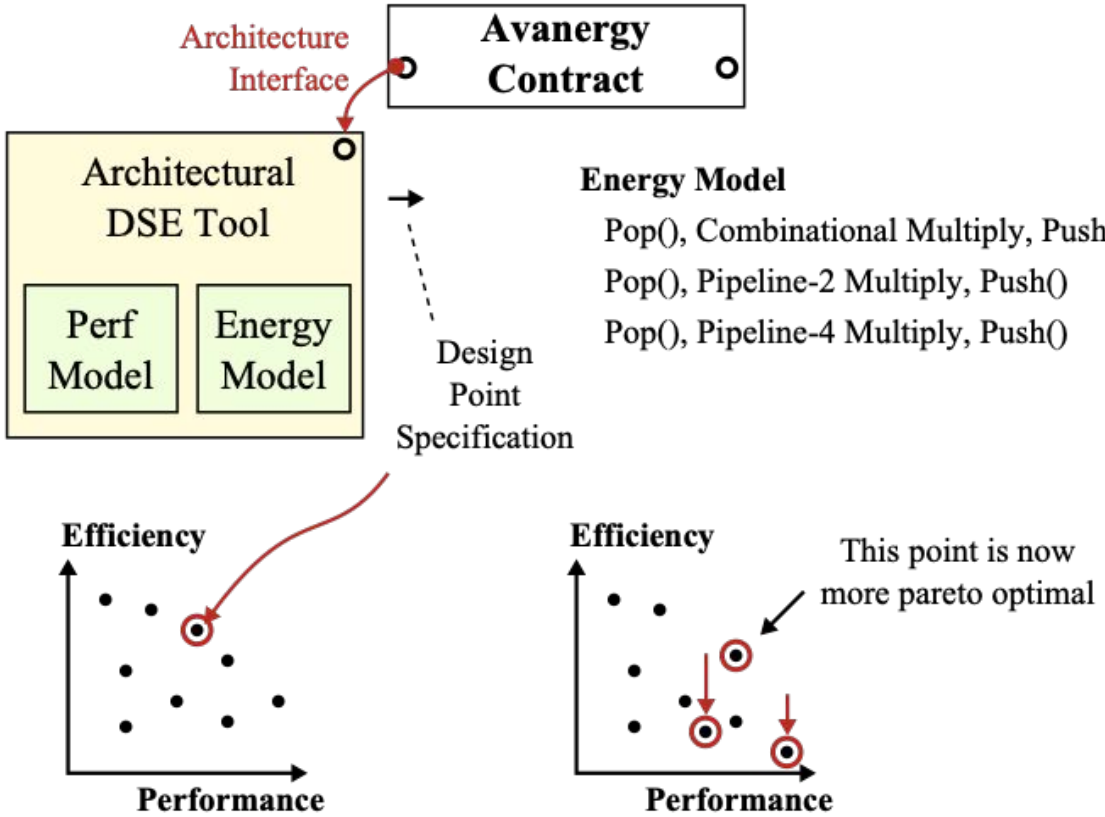
The Avanergy Architecture Interface



Goal of the architecture interface

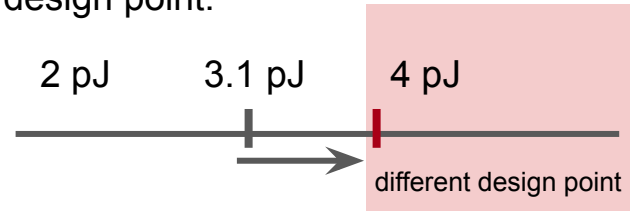
Determine "how wrong" the transaction energy cost must be, where the DSE tool would have instead picked a different design point.

The Avanergy Architecture Interface



Goal of the architecture interface

Determine "how wrong" the transaction energy cost must be, where the DSE tool would have instead picked a different design point.



The Avanergy Architecture Interface

As a recipe:

1. Run many instances of the architectural design-space exploration tool with slightly perturbed values of the energy model
2. Run a relaxation algorithm to discover the most conservative values
3. Deliver energy constraints from these experiments for each entry in the energy model

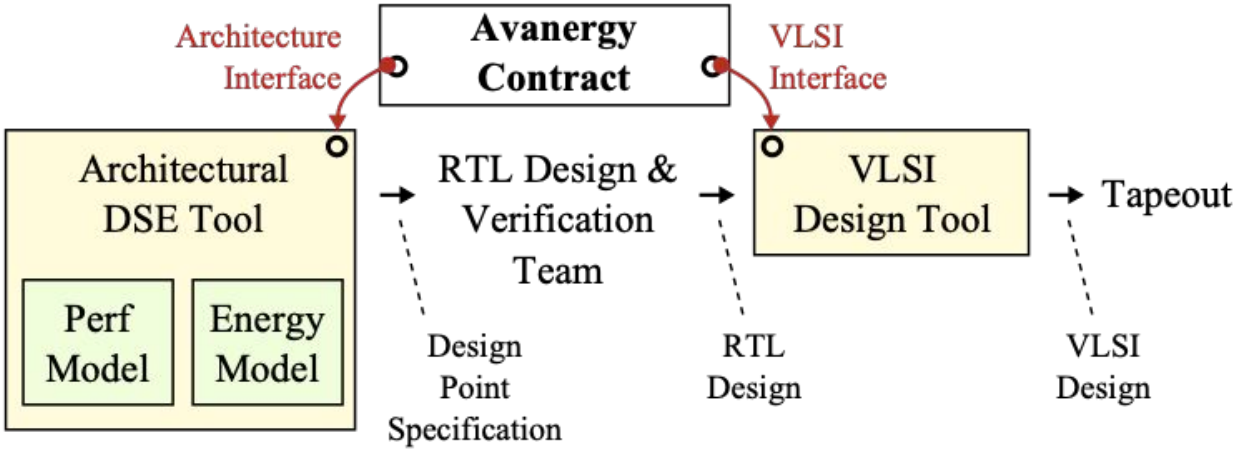
The Avanergy Architecture Interface

As a recipe:

1. Run many instances of the architectural design-space exploration tool with slightly perturbed values of the energy model
2. Run a relaxation algorithm to discover the most conservative values
3. Deliver energy constraints from these experiments for each entry in the energy model

Takeaway: Energy constraints can be purposely constructed with semantics derived from the DSE tool (i.e., the requirements on the energy model to select this design point), and passed down for VLSI to confirm

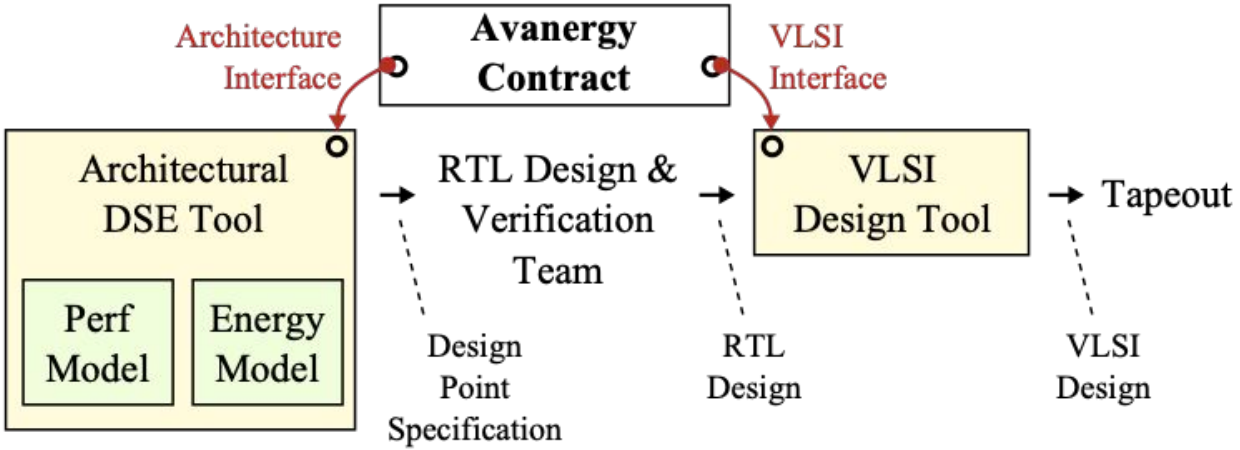
The Avanergy VLSI Interface



Goal of the VLSI interface

Confirm that the energy constraints are met (this type of energy measurement is not a conventional step today)

The Avanergy VLSI Interface



Goal of the VLSI interface

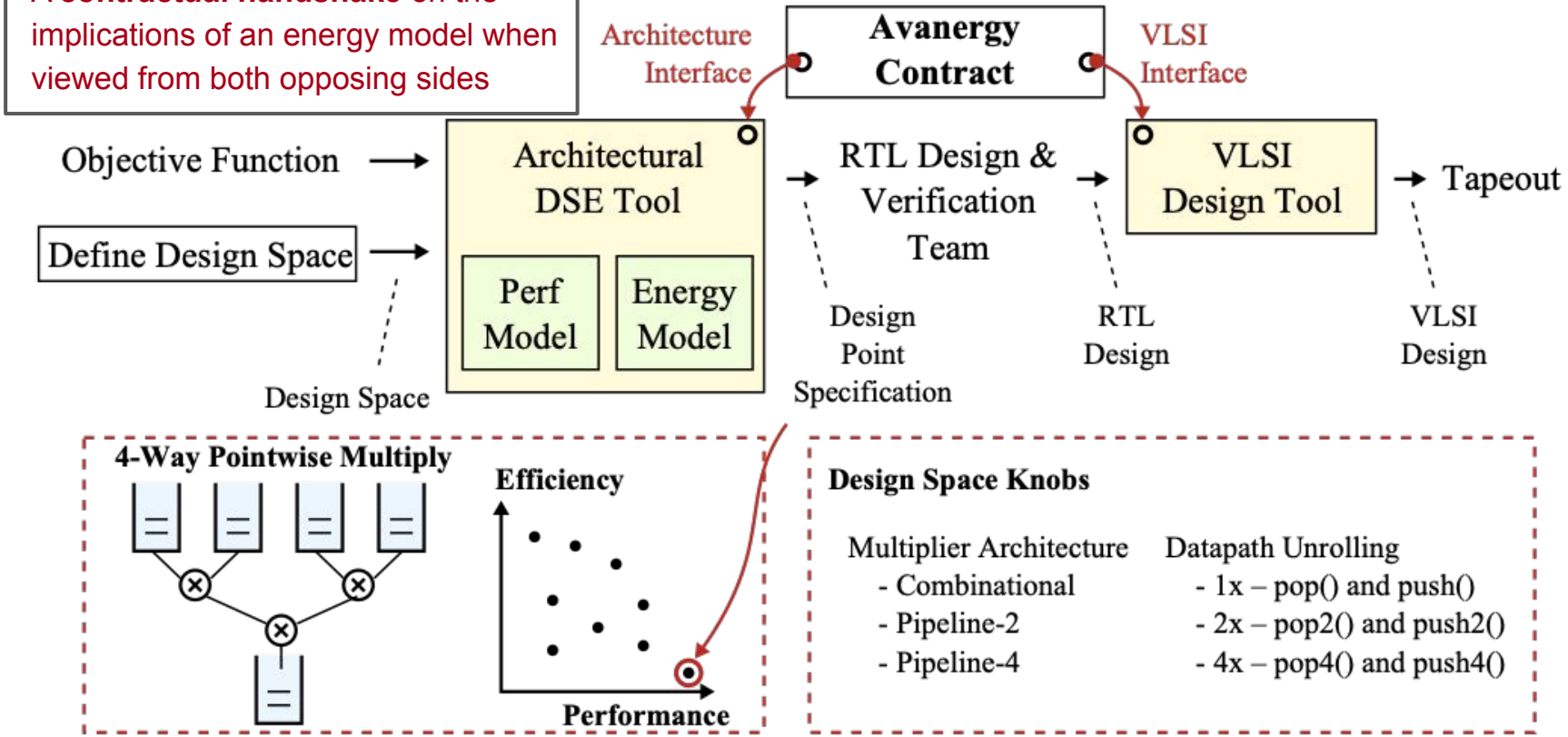
Confirm that the energy constraints are met (this type of energy measurement is not a conventional step today)

```
****************************************
Report : Averaged Power
         -hierarchy
         -nosplit
Design : Tile_MemCore
Version: Q-2019.12-SP2
Date   : Mon Jul  6 19:10:05 2020
****************************************


                                  Int     Switch   Leak     Total
Hierarchy                         Power   Power    Power    Power   %
--------------------------------------------------------------------------------
Tile_MemCore                      5.58e-03 1.90e-03 1.09e-03 8.56e-03 100.0
  WIRE_SB_T3_NORTH_SB_IN_B1 (Tile_MemCore_MuxWrapper_1_1_7)   0.000    0.000    0.000    0.000   N/A
  WIRE_SB_T0_NORTH_SB_IN_B1 (Tile_MemCore_MuxWrapper_1_1_19)  0.000    0.000    0.000    0.000   N/A
  WIRE_SB_T1_SOUTH_SB_IN_B1 (Tile_MemCore_MuxWrapper_1_1_14)  0.000    0.000    0.000    0.000   N/A
  DECODE_FEATURE_0 (Tile_MemCore_Decode08_0)   0.000    0.000 5.29e-07 5.29e-07   0.0
    coreir_eq_8_inst0 (Tile_MemCore_coreir_eq_width8_0)   0.000    0.000 5.29e-07 5.29e-07   0.0
  DECODE_FEATURE_1 (Tile_MemCore_Decode18_0)   0.000    0.000 5.32e-06 5.32e-06   0.1
    coreir_eq_8_inst0 (Tile_MemCore_coreir_eq_width8_321)   0.000    0.000 5.32e-06 5.32e-06   0.1
  DECODE_FEATURE_2 (Tile_MemCore_Decode28_0)   0.000    0.000 6.92e-06 6.92e-06   0.1
    coreir_eq_8_inst0 (Tile_MemCore_coreir_eq_width8_311)   0.000    0.000 6.92e-06 6.92e-06   0.1
  DECODE_FEATURE_3 (Tile_MemCore_Decode38_0)   0.000    0.000 5.96e-11 5.96e-11   0.0
    coreir_eq_8_inst0 (Tile_MemCore_coreir_eq_width8_310)   0.000    0.000 5.96e-11 5.96e-11   0.0
```

Avanergy: An Architecture-VLSI Abstraction (Overview)

A **contractual handshake** on the implications of an energy model when viewed from both opposing sides



Brief Summary of Differences

Implications for architectural design-space exploration framework designers

- In order for these tools to fit into a future of automated, iterative development, your tool must expect to be...
  - Wrapped with an energy model that can be permuted
  - Run many times to discover energy constraints for the first half of this contract

Brief Summary of Differences

Implications for architectural design-space exploration framework designers

- In order for these tools to fit into a future of automated, iterative development, your tool must expect to be...
  - Wrapped with an energy model that can be permuted
  - Run many times to discover energy constraints for the first half of this contract

Implications for VLSI design team engineers

- There is a lack of support for characterizing "transactional" energy
- Future VLSI tools that support automated DSE should expect to establish some form of full-stack handshake

Avanergy: An Architecture-VLSI Abstraction for Automated Energy-Aware Design-Space Exploration Tools

A new world of hardware designers may have (1) more people, (2) operating with less expertise, and (3) relying heavily on automated tools.

Avanergy: An Architecture-VLSI Abstraction for Automated Energy-Aware Design-Space Exploration Tools

A new world of hardware designers may have (1) more people, (2) operating with less expertise, and (3) relying heavily on automated tools.

- Traditional design methodologies are based on the **wrong expertise profiles**

Avanergy: An Architecture-VLSI Abstraction for Automated Energy-Aware Design-Space Exploration Tools

A new world of hardware designers may have (1) more people, (2) operating with less expertise, and (3) relying heavily on automated tools.

- Traditional design methodologies are based on the **wrong expertise profiles**
- A short and **agile iterative approach** can naturally accommodate error

Avanergy: An Architecture-VLSI Abstraction for Automated Energy-Aware Design-Space Exploration Tools

A new world of hardware designers may have (1) more people, (2) operating with less expertise, and (3) relying heavily on automated tools.

- Traditional design methodologies are based on the **wrong expertise profiles**
- A short and **agile iterative approach** can naturally accommodate error
- Automation requires a systematic approach and **quantifiable abstractions**

Avanergy: An Architecture-VLSI Abstraction for Automated Energy-Aware Design-Space Exploration Tools

A new world of hardware designers may have (1) more people, (2) operating with less expertise, and (3) relying heavily on automated tools.

- Traditional design methodologies are based on the **wrong expertise profiles**
- A short and **agile iterative approach** can naturally accommodate error
- Automation requires a systematic approach and **quantifiable abstractions**

Avanergy is a new abstraction that allows a quantitative and systematic view of the architecture-VLSI boundary to allow energy-aware design-space exploration in a way that future automated tools can build upon it.

BACKUP