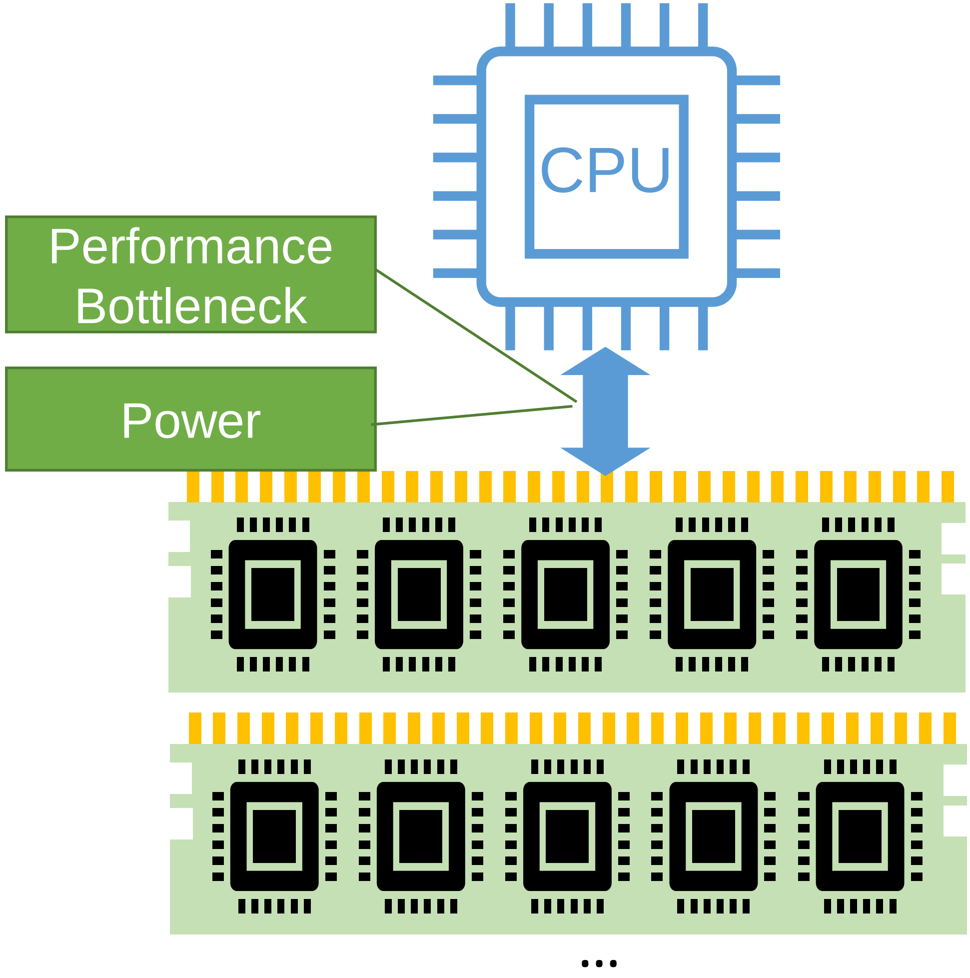
# SecNDP: Secure Near-Data Processing with Untrusted Memory

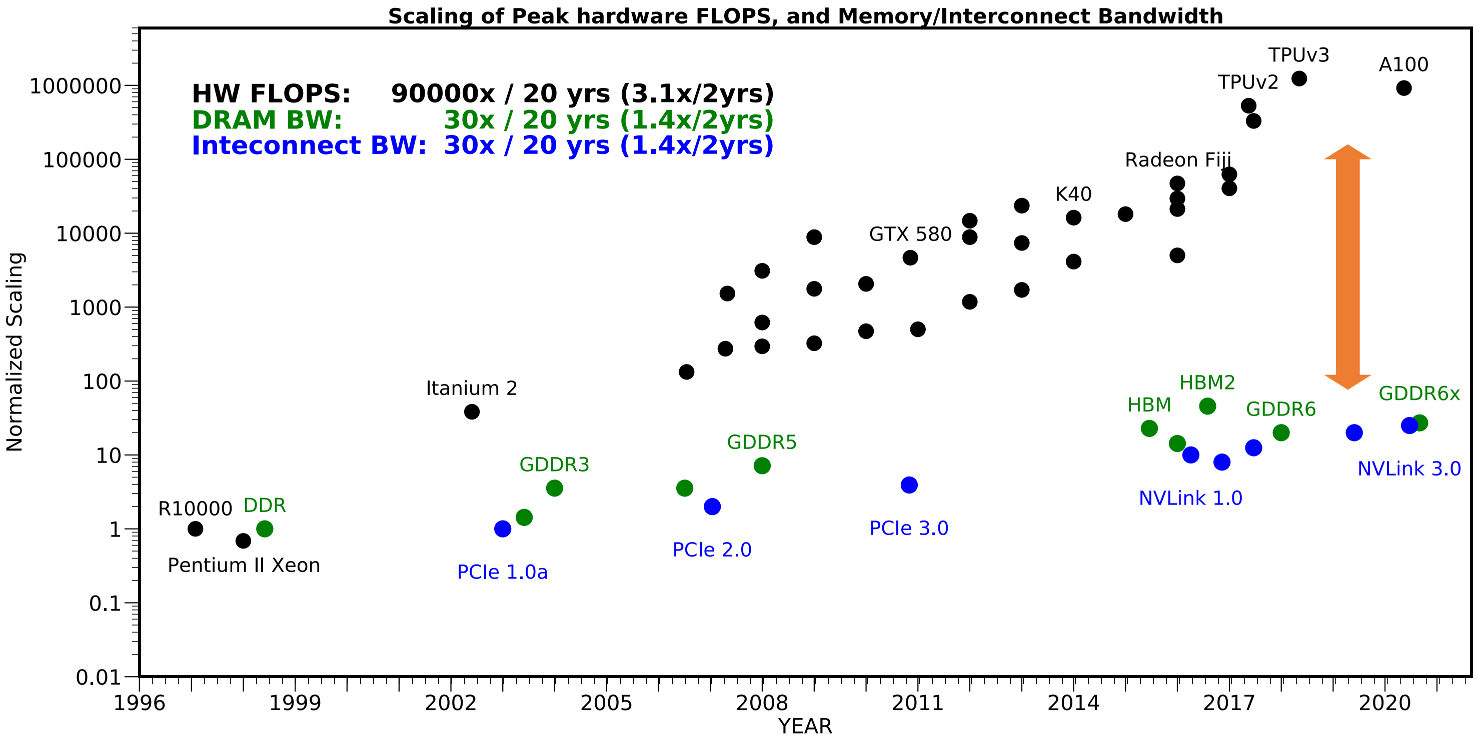
Wenjie Xiong[*§], Liu Ke[*†§], Dimitrije Jankov[‡], Michael Kounavis[*],
Xiaochen Wang[*], Eric Northup[*], Jie Amy Yang[*],
Bilge Acun[*], Carole-Jean Wu[*], Ping Tak Peter Tang[*],
G. Edward Suh[*◇], Xuan Zhang[†], Hsien-Hsin S. Lee[*]

[*]Meta, [†]Washington University in St. Louis,
[‡]Rice University, [◇]Cornell University

[§]The first two authors contributed equally to this work

# “Memory Wall”

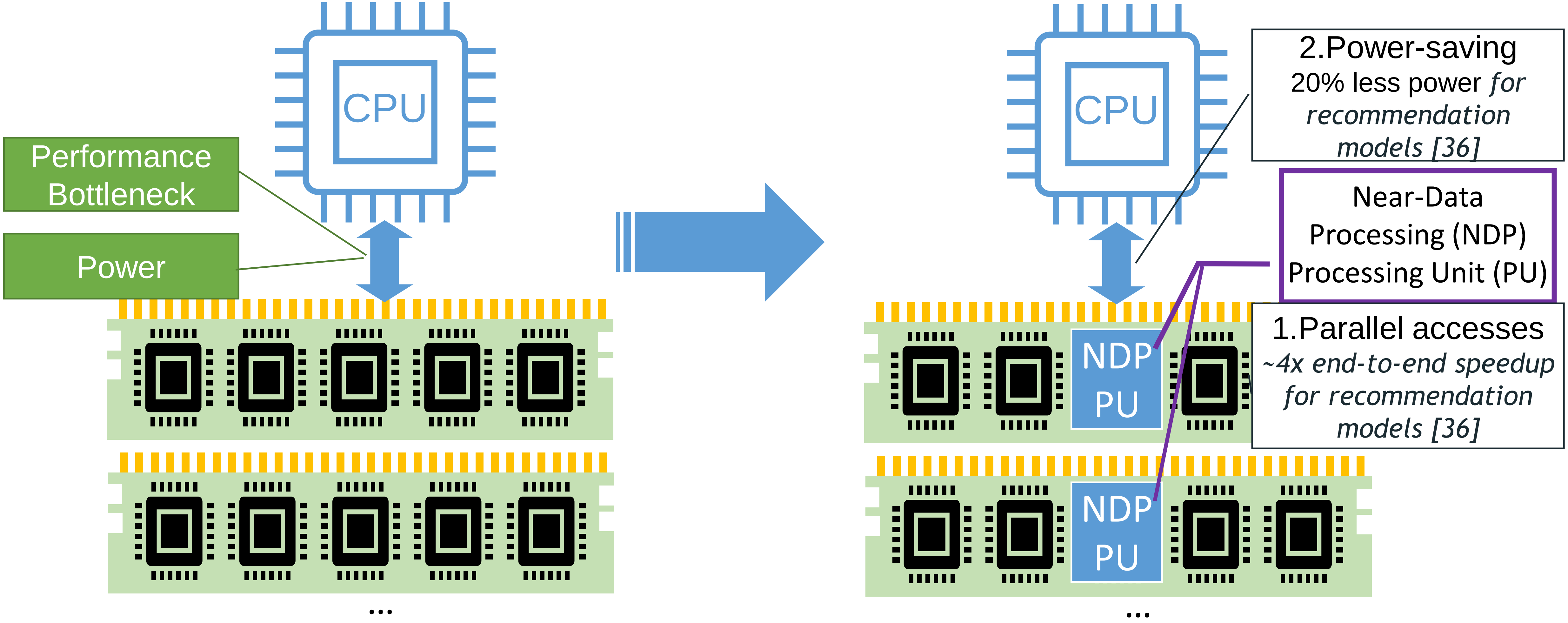



The scaling of the bandwidth of interconnections and memory, as well as the Peak FLOPS. As can be seen, the bandwidth is increasing very slowly.

Source: https://medium.com/riselab/ai-and-memory-wall-2cb4265cb0b8

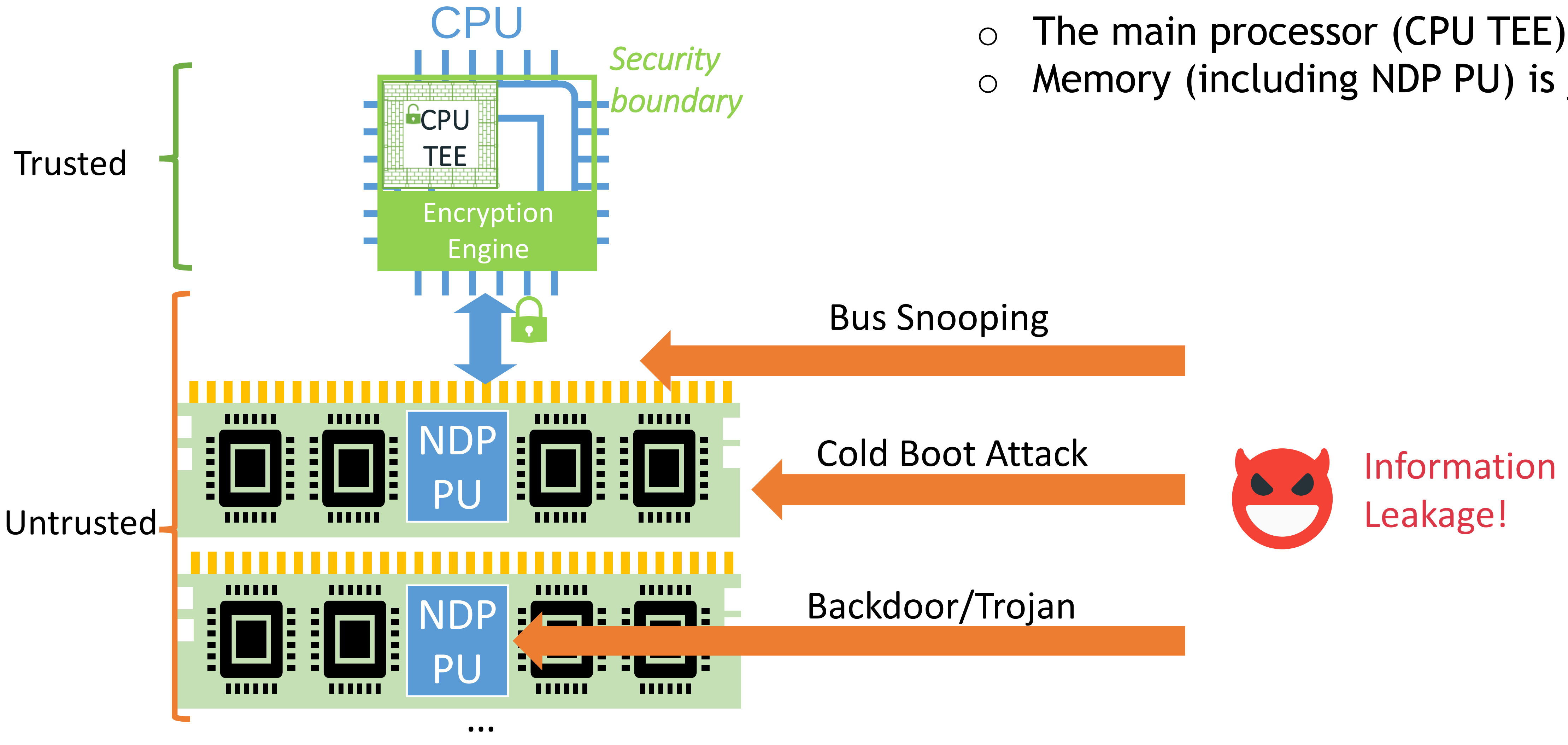
# “Memory Wall” and Near Data Processing (NDP)



NDP Example

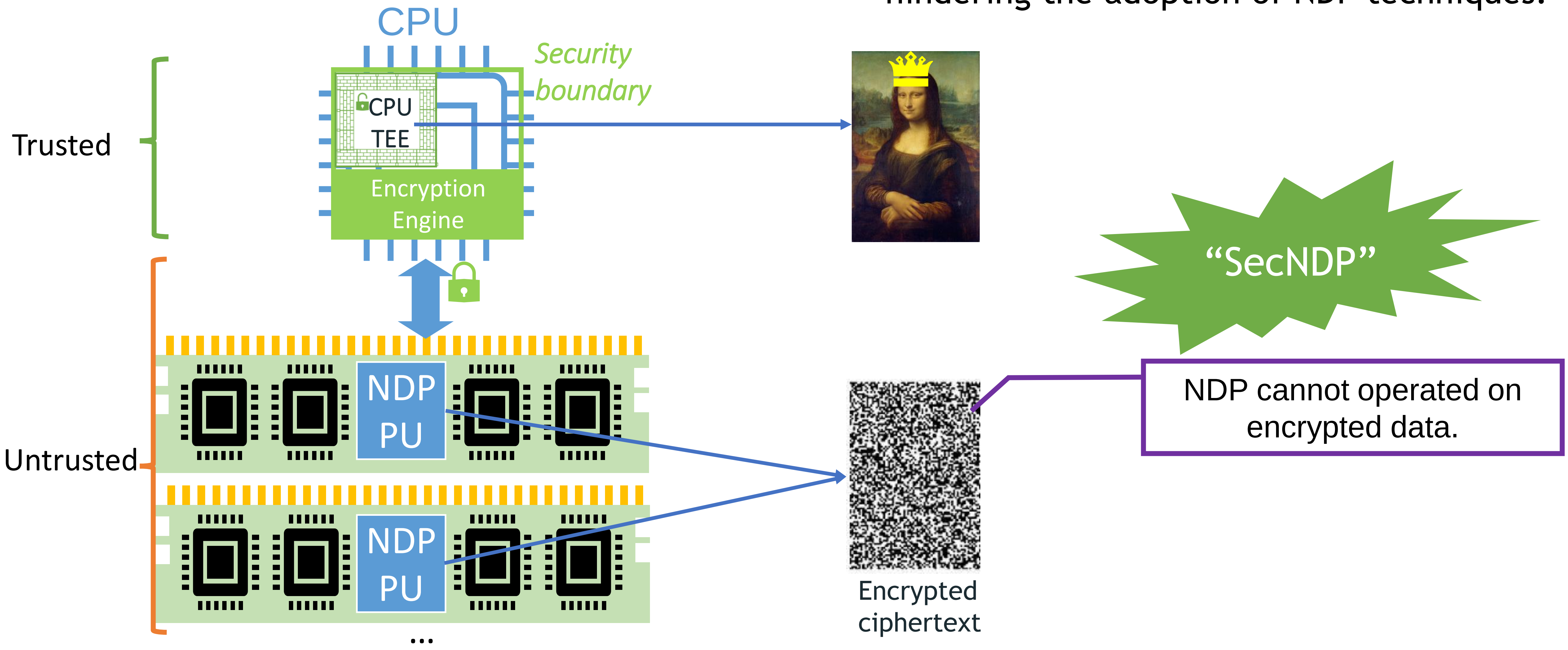
# CPU TEEs Today

- **Trusted Execution Environment (TEE)** provides strong *confidentiality* and *integrity* protection using hardware.
- Threat model:
  - The main processor (CPU TEE) is trusted.
  - Memory (including NDP PU) is untrusted.

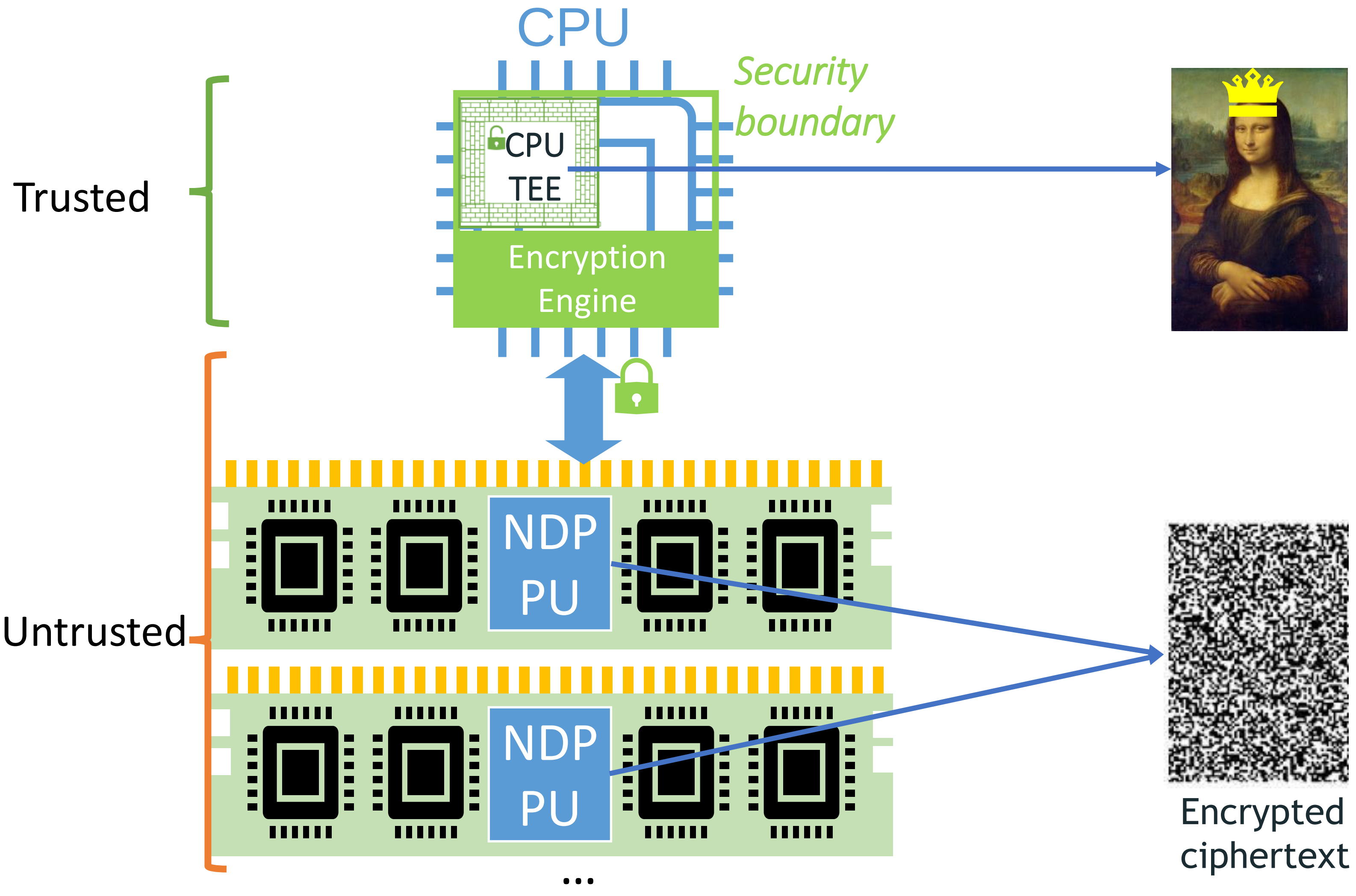


# CPU TEEs and Challenges

- Off-chip data is protected by encryption.
- Current memory encryption <u>does not support computation over ciphertext</u>, hindering the adoption of NDP techniques.



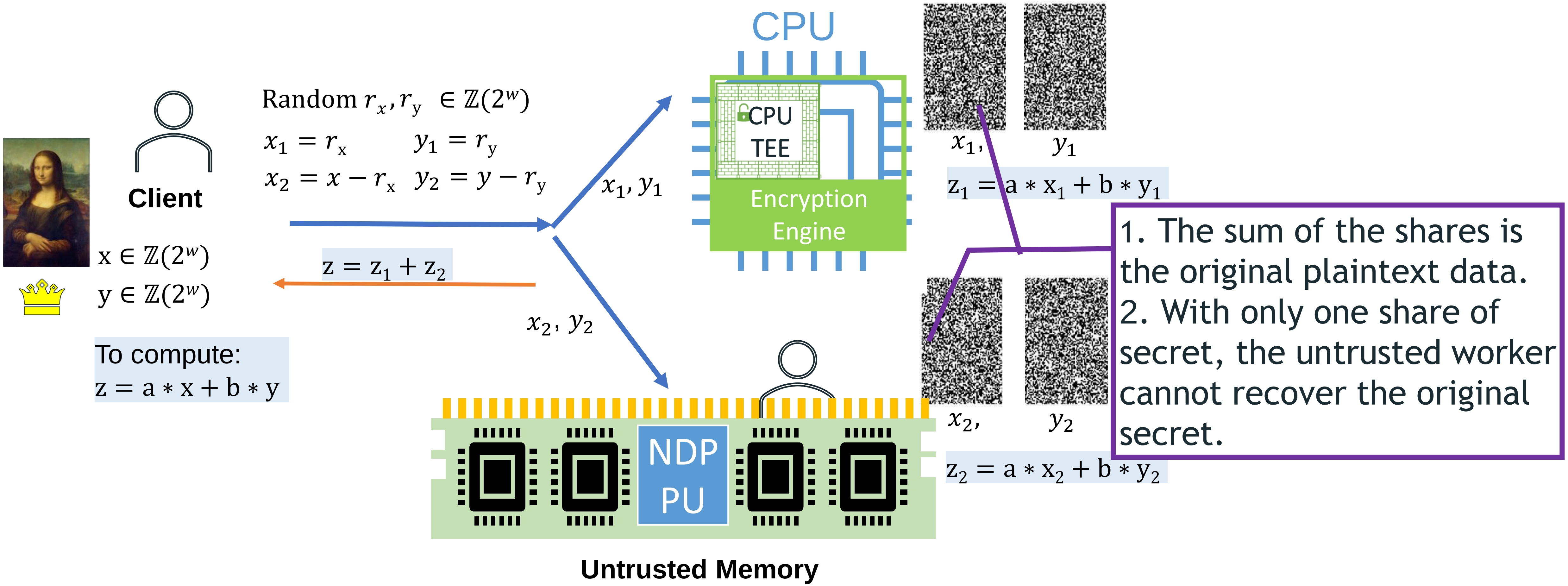
# Contributions of SecNDP




- An <u>encryption scheme</u> that allows computation over ciphertext in the untrusted NDP.
- An <u>integrity verification</u> scheme to validate the correctness of linear operations in NDP.
- Demonstrate performance approaching to that of unprotected NDP.

High-bandwidth, low-power, secure near-data processing!

# Background: Arithmetic Secret Sharing in secure Multi-Party Computation




MPC protects confidentiality from an untrusted party (untrusted NDP).

However, existing MPC scheme assume each party will use the same amount of computing resources and memory.

# SecNDP Encryption

**Encryption** (in the processor)

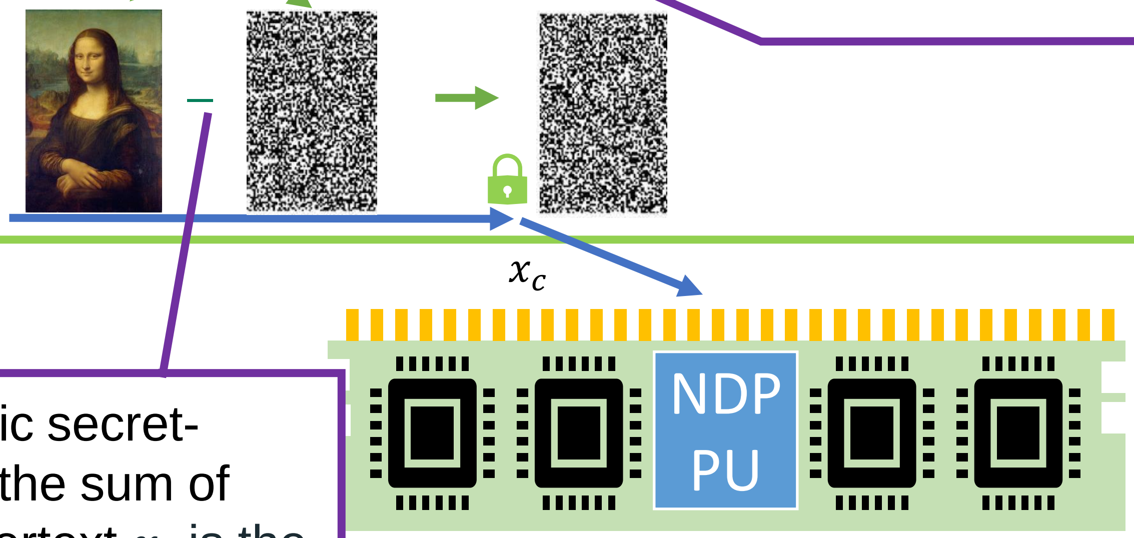
a.k.a., "One time pad (OTP)" or "encrypted counter" using a block cipher

$x_e = Enc_K(addr, version)$

$\boldsymbol{x_c} = \boldsymbol{x} - \boldsymbol{x_e}$

**Plaintext data**

$\mathbf{x} \in \mathbb{Z}(\mathbf{2}^{w})$

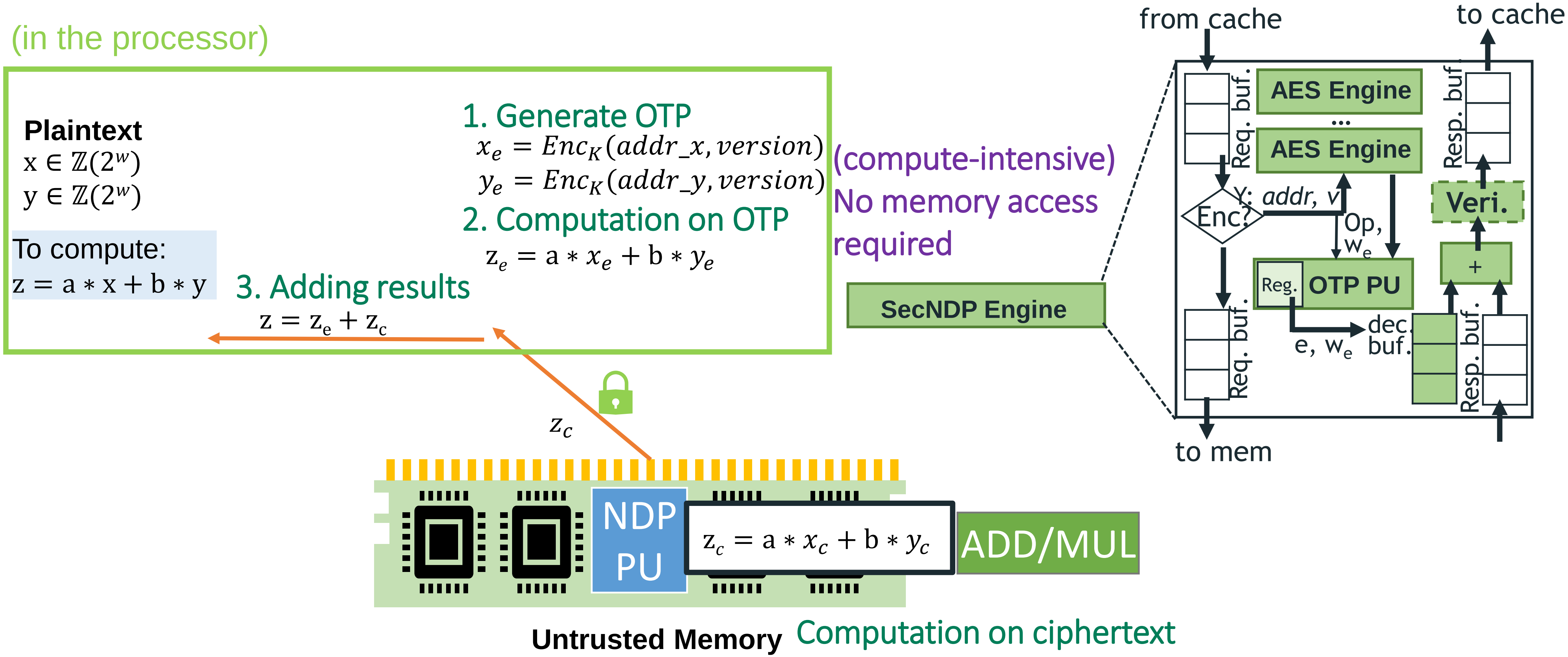



**Untrusted Memory**

OTP takes the address and a version number as inputs. On-chip AES engines can generate OTP efficiently in parallel with memory accesses.

Like the arithmetic secret-sharing in MPC, the sum of OTP $\boldsymbol{x_e}$ and ciphertext $\boldsymbol{x_c}$ is the plaintext data.
We can do computation over $\boldsymbol{x_c}$!

# Computation using SecNDP




Computation on ciphertext
(memory-intensive)

- Parallel accesses
- Save memory traffic

- Integrity Protection: A linear hash function is used to verify the result of linear operations.
- Formally prove the security of the schemes.

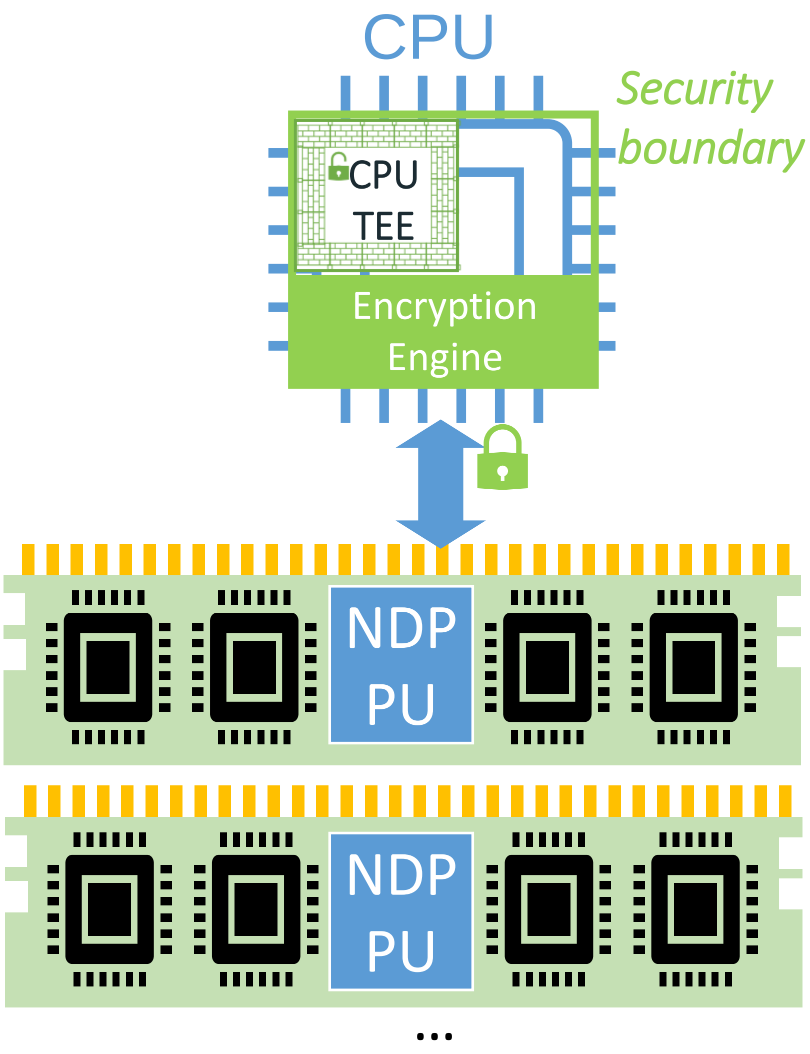
# Performance Evaluation of SecNDP

- Workloads:
  - Deep Learning Recommendation Model (RMC in the table) [58]
    - Embedding table lookups: memory-intensive, executed in NDP
    - Fully connected layers: compute-intensive, executed in the CPU TEE
  - Medical Data analytics, e.g., sum: memory-intensive, executed in NDP
- With enough AES engines, SecNDP's system performance is close to unprotected NDP while providing data confidentiality and integrity guarantee.

End-to-end speedup, with 8 NDP PUs.

| | | RMC1-small | RMC1-large | RMC2-small | RMC2-large | Data Analytics |
|---|---|---|---|---|---|---|
| | unprotected non-NDP | 1x | 1x | 1x | 1x | 1x |
| | unprotected NDP | 2.46x | 3.11x | 4.05x | 4.44x | 7.46x |
| | SecNDP | 2.36x | 3.02x | 3.95x | 4.33x | 7.46x |
| Existing CPU TEE | SGX-CFL | 0.0038x | 0.0037x | N/A | N/A | 0.1738x |
| Existing CPU TEE | SGX-ICL (no int. tree) | 0.59x | 0.60x | N/A | N/A | 0.57x |

# Conclusions




SecNDP is the first work to enable a TEE to leverage the performance and energy benefits of untrusted NDP securely.

- We proposed an encryption scheme allow computation over ciphertext in the untrusted NDP.
- We proposed an integrity verification scheme to validate the correctness of the computation in NDP.

SecNDP schemes demonstrate performance approaching to that of unprotected NDP.

- Performance (7.46x speedup)
- Energy consumption (18% energy-saving)
- Accuracy (negligible impact)

Thank you!