Input

- Workload
- Hardware
- SLAs
- **Performance Constraints**
  - Initial design
  - Time threshold
  - Distance to optimal

Learned Shortcuts

High confidence

Low confidence

Feedback

Partial design

Solved

Currently solving

Not yet solved

?

Design Search

Update design

Iterative Search

Design Continuums

<u>Fast</u> analytical model optimisation

$$\arg\min f(x)$$

Is it a design continuum?

Node by node design process

Systematically evaluate various designs for each node

Otherwise

Memoization

H/W Pruning

Expert Rules

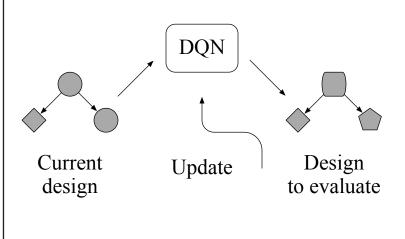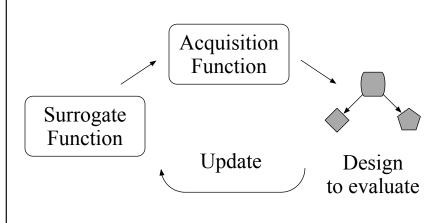parallelization

Reinforcement Learning

feedback

Neural Net

Cost Estimation/ Synthesis

Deep Reinforcement Learning

Bayesian Optimization

Genetic Algorithms

Output

Data structure design

AST

C++ Code

Data Layout and Index Synthesis

Key Value Data Structures

$\langle k, v \rangle$

Layout Primitives

Data Node (Element) Design Space

- Node has partitioning?
- Node has Bloom filters?
- ...
- Node has Zone maps?

Partitioning

Bloom Filter

Zone Maps

True False

Min Max Both

Func. 1 Func. 2

primitives

Data Node (Element) Library

- Linked List
- B+Tree Internal
- B-Tree Internal
- Design Continuums
- Trie
- Skip List
- Data Page
- Array

Overall Designs

Design Space

Generalized Cost and Algorithm Synthesis

Data Access Primitives

Level 1

Level 2

- Serial Scan
- Equality Scan
- Range Scan
- Sorted Search
- Binary Search
- Random Probe

C++

C++

C++

Micro-benchmarks

Cost Synthesizer

Machine Learning

Micro-benchmarks train models on different hardware profiles.

$$f(x) = \sum_i \frac{c_i}{1 + e^{-k_i(x - x_i)}}$$

$$f(x) = ax + b$$

Hardware Profiles

NVM
GB
GB
GFX

Equality Scan

Range Scan

Binary Search

Code Generator

Operation Synthesis

Operation Synthesis (Level 1)

Hardware Conscious Synthesis (Level 2)

Put **Get** **Delete** **Bulk Load**

Level 1 to Level 2 translation

Put **Get** **Delete** **Bulk Load**

Concurrency & Updates

Space Efficiency Optimization

Foreground

Background

Image

Segmentation

$$\begin{bmatrix} 0.1 \dots 1.2 \\ \dots \\ \dots \\ 1.8 \dots 7.6 \end{bmatrix}$$

$$\begin{bmatrix} 1.7 \dots 2.2 \\ \dots \\ \dots \\ 4.8 \dots 1.3 \end{bmatrix}$$

Transform

Quantization

$$\begin{bmatrix} 0 \dots 1 \\ \dots \\ \dots \\ 2 \dots 8 \end{bmatrix}$$

$$\begin{bmatrix} 2 \dots 2 \\ \dots \\ \dots \\ 5 \dots 1 \end{bmatrix}$$

Coding

Dictionary Compression

Repeat

D1   D2

D2   D3

D3

Current
popn.

Evaluate and
select

Recombine
and mutate

D2

D2'

D3'

Genetic Algorithms

DQN

Current
design

Update

Design
to evaluate

Deep Reinforcement Learning

Acquisition
Function

Surrogate
Function

Update

Design
to evaluate

Bayesian optimization