# ChipXplore: Natural Language Exploration of Hardware Designs and Libraries

Manar Abdelatty, Jacob Rosenstein, Sherief Reda

*Brown University, Providence, RI*

# Manual Navigation of PDK and Design Files

- Engineers typically navigate PDK files during various stages of hardware design.

# Navigating PDKs is Time Consuming and Error Prone

Process Corners

Min | Nom | Max

**Technology Lef**

```
LAYER met1
  TYPE ROUTING ;
  DIRECTION HORIZONTAL ;

  PITCH 0.34 ;
  OFFSET 0.17 ;

  WIDTH 0.14 ;                # Met1 1
  # SPACING 0.14 ;             # Met1 2
  # SPACING 0.28 RANGE 3.001 100 ; # Met1 3b
  SPACINGTABLE
    PARALLELRUNLENGTH 0
    WIDTH 0 0.14
    WIDTH 3 0.28 ;
  AREA 0.083 ;                # Met1 6
  THICKNESS 0.35 ;
  MINENCLOSEDAREA 0.14 ;

  ANTENNAMODEL OXIDE1 ;
  ANTENNADIFFSIDEAREARATIO PWL ( ( 0 400 ) ( 0.0125 400 ) ... ;

  EDGECAPACITANCE 40.567E-6 ;
  CAPACITANCE CPERSQDIST 25.7784E-6 ;
  DCCURRENTDENSITY AVERAGE 2.8 ; # mA/um Iavg_max at Tj = 90oC
  ACCURRENTDENSITY RMS 6.1 ; # mA/um Irms_max at Tj = 90oC
  MAXIMUMDENSITY 70 ;
  DENSITYCHECKWINDOW 700 700 ;
  DENSITYCHECKSTEP 70 ;

  RESISTANCE RPERSQ 0.125 ;
END met1
```

Different Standard Cell Libraries

- High Speed
- Low Power
- …
- High Density

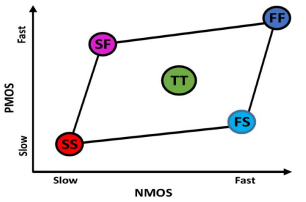
**Lef**

```
MACRO a2bb2o_1
  CLASS CORE ;
  FOREIGN a2bb2o_1 ;
  ORIGIN 0.000 0.000 ;
  SIZE 3.680 BY 2.720 ;
  SYMMETRY X Y R90 ;
  SITE unithd ;
  PIN A1_N
    DIRECTION INPUT ;
    USE SIGNAL ;
    ANTENNAGATEAREA 0.126000 ;
    PORT
      LAYER li1 ;
        RECT 0.910 0.995 1.240 1.615 ;
    END
  END A1_N
  ...
  PIN X
    DIRECTION OUTPUT ;
    USE SIGNAL ;
    ANTENNADIFFAREA 0.429000 ;
    ...
  END X
  OBS
      LAYER li1 ;
        RECT 0.000 2.635 3.680 2.805 ;
        PORT
        ...
        END
  END
END a2bb2o_1
```

**Liberty**

```
library ("tt_025C_1v80") {
..
cell ("a41o_1") {
   area : 10.009600000;
   cell_footprint :  "a41o";
   cell_leakage_power : 0.0066084930;
   driver_waveform_fall : "ramp";
   driver_waveform_rise : "ramp";
   pin ("A1") {
      capacitance : 0.0023100000;
      clock : "false";
      direction : "input";
      fall_capacitance : 0.0022460000;
      max_transition : 1.5000000000;
      ...
      rise_capacitance : 0.0023740000;
   }
   ...
   pin ("X") {
      direction : "output";
      function : "(A1&A2&A3&A4) | (B1)";
      timing () {
         cell_fall ("del_1_7_7") {
            index_1("0.0100000000, 0.0230506000, ...");
            index_2("0.0005000000, 0.0013153400,, ...");
            values("0.1409252000, 0.1475898000, ...);
         }
         ....
      }
   }
}
```

At Different PVT Corners



# Navigate PDKs with Natural Language

What is the area of the *and2:1* cell in the high speed library?

The *and2:1* cell has an area of 12.0 units

**Retrieve**

How can LLMs retrieve from the PDK files ?

**PDK Files**

**Liberty**

```
library ("tt_025C_1v80") {
..
cell ("a41o_1") {
   area : 10.009600000;
   cell_footprint :  "a41o";
   cell_leakage_power : 0.0066084930;
   driver_waveform_fall : "ramp";
   driver_waveform_rise : "ramp";
   pin ("A1") {
      capacitance : 0.0023100000;
      clock : "false";
      direction : "input";
      fall_capacitance : 0.0
      max_transition : 1.500
      ...
      rise_capacitance : 0.00
   }
   ...
   pin ("X") {
      direction : "output";
      function : "(A1&A2&A3&A4) ... B1)";
      timing () {
         cell_fall ("del_1_7_7") {
            index_1("0.0100000000, 0.0230506000, ...");
            index_2("0.0005000000, 0.0013153400,, ...");
            values("0.1409252000, 0.1475898000, ...");
         }
         ....
      }
   }
}
```

**Lef**

```
LAYER met1
  TYPE ROUTING ;
  DIRECTION HORIZONTAL ;

  PITCH 0.34 ;
  OFFSET 0.17 ;

  WIDTH 0
                                  t1 3b
                                  # Met1 6
  MINENCLOSEDAREA 0.14 ;

  ANTENNAMODEL OXIDE1 ;
  ANTENNADIFFSIDEAREARATIO PWL ( ( 0 400 ) ( 0.0125 400 ) ... ;

  EDGECAPACITANCE 40.567E-6 ;
  CAPACITANCE CPERSQDIST 25.7784E-6 ;
  DCCURRENTDENSITY AVERAGE 2.8 ; # mA/um Iavg_max at Tj = 90oC
  ACCURRENTDENSITY RMS 6.1 ; # mA/um Irms_max at Tj = 90oC
  MAXIMUMDENSITY 70 ;
  DENSITYCHECKWINDOW 700 700 ;
  DENSITYCHECKSTEP 70 ;

  RESISTANCE RPERSQ 0.125 ;
END met1
```

Exploit the Structured Property


```
EDGECAPACITANCE 40.567E-6 ;
CAPACITANCE CPERSQDIST 25.7784E-6 ;
DCCURRENTDENSITY AVERAGE 2.8 ; # mA/um Iavg_max at Tj = 90oC
ACCURRENTDENSITY RMS 6.1 ; # mA/um Irms_max at Tj = 90oC
MAXIMUMDENSITY 70 ;
DENSITYCHECKWINDOW 700 700 ;
DENSITYCHECKSTEP 70 ;

RESISTANCE RPERSQ 0.125 ;
END met1
```

# PDK: Relational Database

**PDK**

Relational Database

**RoutingLayers**

| LayerID | Name | Type | Direction | ResistancePerSQ | ... | Corner | Standard_Cell_Library |
|---|---|---|---|---|---|---|---|

**AntennaDiffSideAreaRatios**

| RatioID | LayerID | Type | X1 | Y1 | X2 | Y2 |
|---|---|---|---|---|---|---|

**CutLayers**

| LayerID | Name | Type | Width | Spacing | Resistance | ... | Corner | Standard_Cell_Library |
|---|---|---|---|---|---|---|---|---|

**AntennaDiffAreaRatios**

| RatioID | LayerID | Type | X1 | Y1 | X2 | Y2 |
|---|---|---|---|---|---|---|

**Vias**

| ViaID | Name | Lower_Layer | Upper_Layer | Corner | Standard_Cell_Library |
|---|---|---|---|---|---|

**ViaLayers**

| ViaLayerID | ViaID | LayerName | RectX1 | RectY1 | RectX2 | RectY2 |
|---|---|---|---|---|---|---|

**Technology Lef**

**Macros**

| MacroID | Name | Class | ForeignName | SizeWidth | ... | Standard_Cell_Library |
|---|---|---|---|---|---|---|

**Pins**

| PinID | MacroID | Name | Direction | Use | AntennaDiffArea | AntennaGateArea |
|---|---|---|---|---|---|---|

**PinPorts**

| PortID | PinID | Layer |
|---|---|---|

**PinPortRectangles**

| RectID | PortID | RectX1 | RectY1 | RectX2 | RectY2 |
|---|---|---|---|---|---|

**Obstructions**

| ObstructionID | MacroID | Layer |
|---|---|---|

**ObstructionRectangles**

| ObstructionRectID | ObstructionID | RectX1 | RectY1 | RectX2 | RectY2 |
|---|---|---|---|---|---|

**LEF**

**OperatingConditions**

| Condition_ID | Name | Voltage | Temperature | Process | ... | Standard_Cell_Library |
|---|---|---|---|---|---|---|

**Cells**

| Cell_ID | Name | Area | Leakage_Power | Footprint | Waveform_Fall | ... | Condition_ID |
|---|---|---|---|---|---|---|---|

**InputPins**

| Input_Pin_ID | Cell_ID | Name | Clock | Capacitance | ... | Rise_Capacitance |
|---|---|---|---|---|---|---|

**OutputPins**

| Output_Pin_ID | Cell_ID | Name | Function | Max_Transition | ... | Max_Capacitance |
|---|---|---|---|---|---|---|

**TimingValues**

| Timing_Value__ID | Output_Pin_ID | Related_Input_Pin | ... | Fall_Delay |
|---|---|---|---|---|

**Liberty**

# Hardware Design (DEF): Graph Database

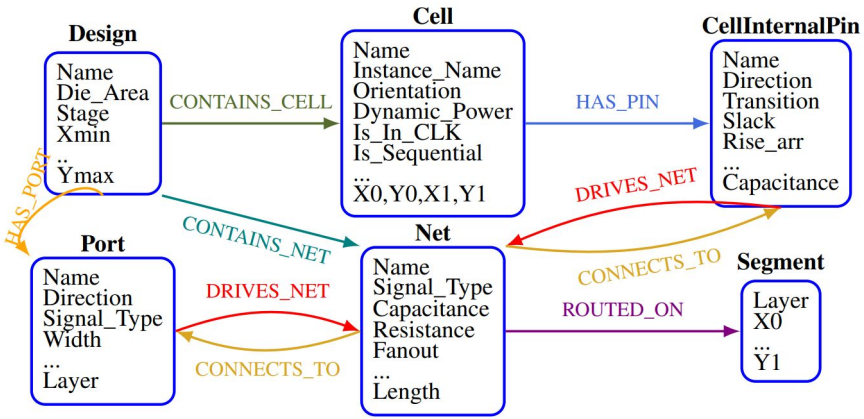
```
DESIGN mydesign ;
UNITS DISTANCE MICRONS 1000 ;
DIEAREA ( 0 0 ) ( 96990 107710 ) ;
COMPONENTS 391 ;
   - _102_ clkbuf_4 + PLACED ( 31740 51680 ) FS ;
   - _103_ a21o_1 + PLACED ( 63940 73440 ) S ;
   ...
END COMPONENTS
PINS 38 ;
   - a[0] + NET a[0] + DIRECTION INPUT + USE SIGNAL
    + PORT
      + LAYER met2 ( -140 -2000 ) ( 140 2000 )
      + PLACED ( 5750 105710 ) N ;
   - a[10] + NET a[10] + DIRECTION INPUT + USE SIGNAL
    + PORT
      + LAYER met2 ( -140 -2000 ) ( 140 2000 )
      + PLACED ( 33350 105710 ) N ;
   ...
END PINS
NETS 265 ;
   - _000_ ( _240_ A ) ( _235_ A2 )  .. + USE SIGNAL ;
   - _016_ ( _250_ B ) ( _249_ Y ) .. + USE SIGNAL ;
   ...
END NETS
END DESIGN
```

DEF

Desig

n

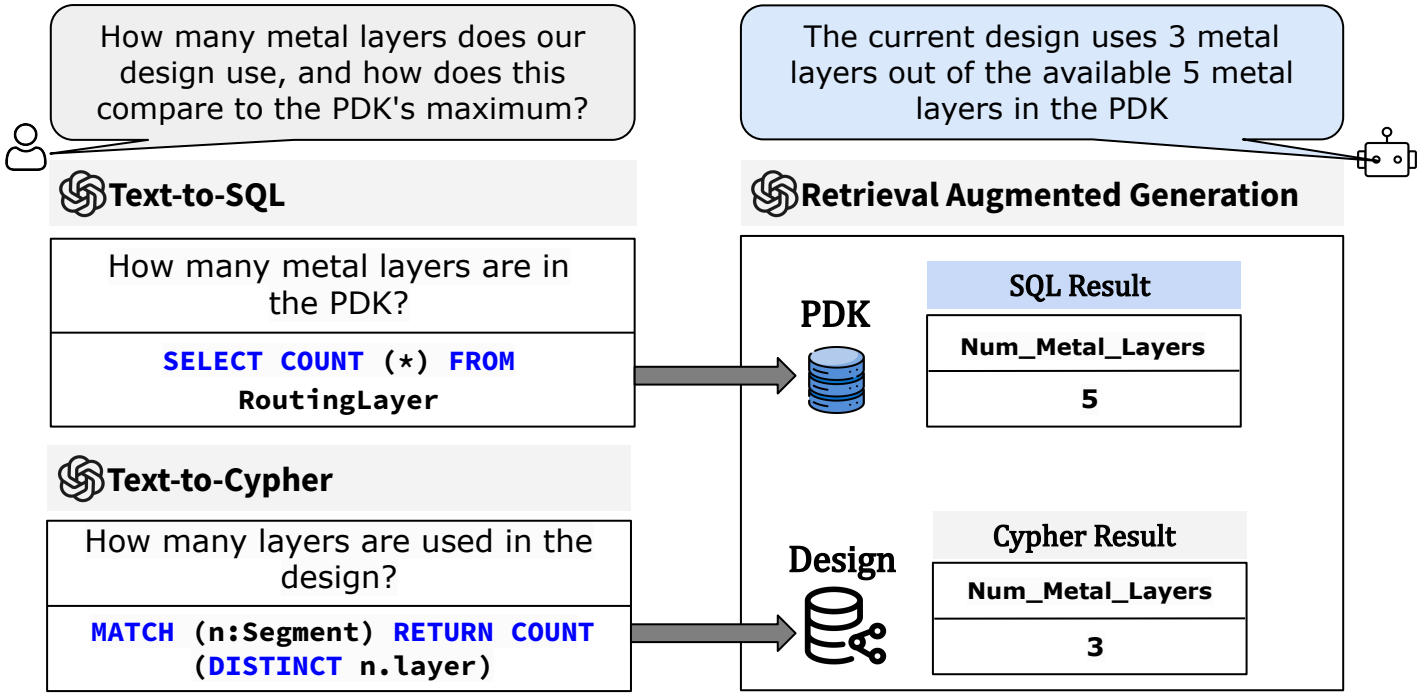
Graph Database

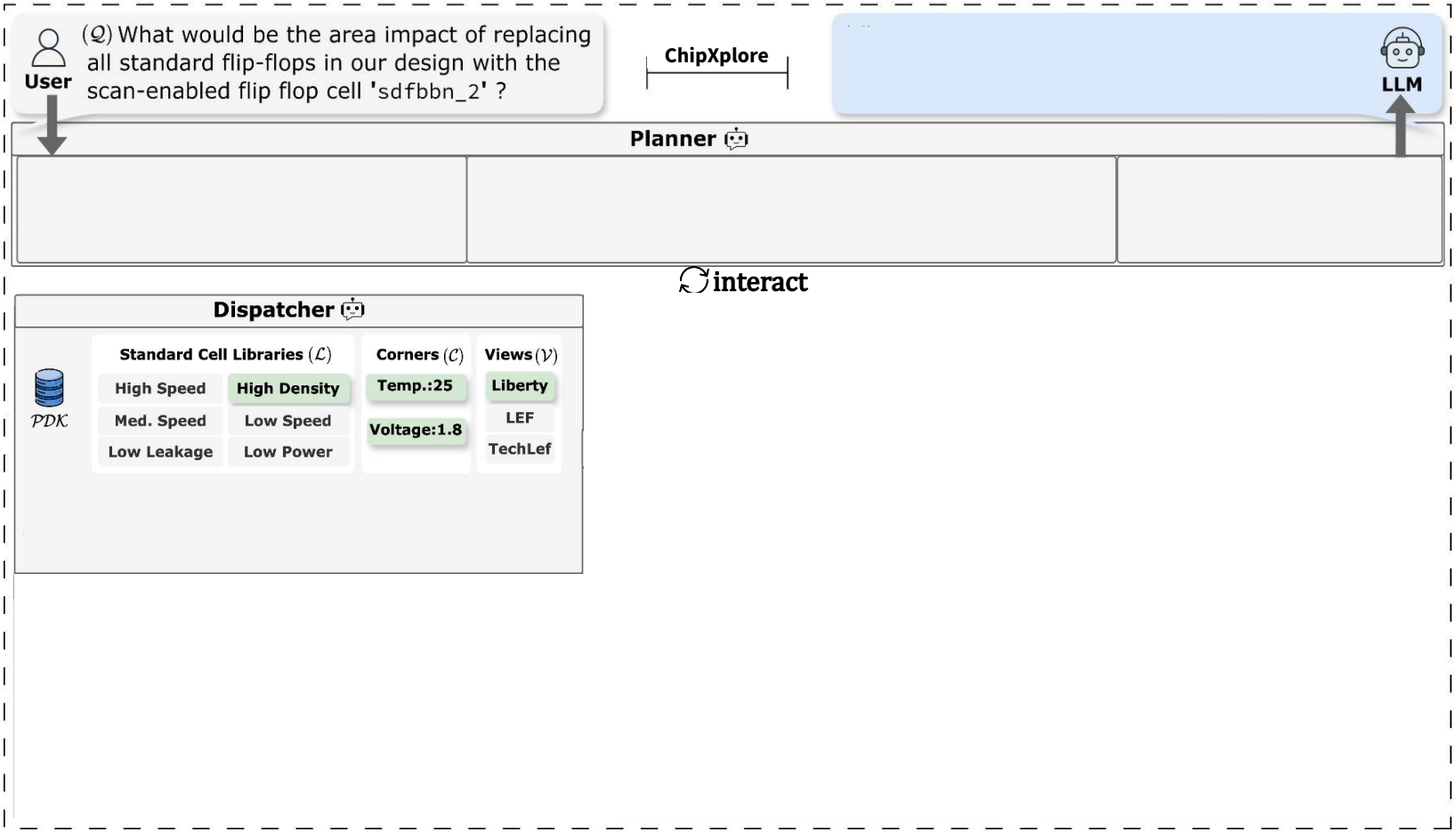



DEF

# LLM-based Retrieval with Database Queries

- Dynamically generate database queries based on user question.






ChipXplore Framework

# Experimental Results: Setup

→ **LLMs**
- gpt-4-turbo
- Llama 3.3:70b
- DeepSeek-V3

→ **PDK**: Skywater 130nm PDK

**PDK Database**

| | |
|---|---|
| **Standard Cell Libraries (SCLs)** | 6 |
| **Tables** | 19 |
| **Cell Entries** | 39,576 |
| **Timing Entries** | 13,874,290 |
| **Size** | 1.1GB |

→ **Design**: Picorv CPU design.

**DEF Design**

| | |
|---|---|
| **Picorv CPU Design - Cells** | 50,637 |
| **Picorv CPU Design - Nets** | 11,070 |

**Neo4j Database**

| | |
|---|---|
| **Picorv CPU Neo4j DB- Nodes** | 605,563 |
| **Picorv CPU Neo4j DB- Edges** | 2,582,105 |

# Use Cases: Cross Library Comparison

**(a) User Question**

Compare the leakage power, clock pin capacitance, area, and average propagation delay of the flip-flop cells with drive strength 4 across the different libraries in the PDK

**Generated SQL**

```
SELECT OC.Cell_Library, C.Name, C.Cell_Leakage_Power, C.Area,
IP.Clock_Pin_Capacitance, TV.Avg_Propagation_Delay FROM
Operating_Conditions OC JOIN Cells C ON OC.Condition_ID =
C.Condition_ID LEFT JOIN Input_Pins IP ON C.Cell_ID = IP.Cell_ID AND
IP.Clock = True LEFT JOIN (SELECT Cell_ID, AVG(Average_Delay) AS
Avg_Propagation_Delay FROM Timing_Values GROUP BY Cell_ID) TV ON
C.Cell_ID = TV.Cell_ID WHERE OC.Temperature = 25.0 AND
OC.Voltage = 1.8 AND C.Is_Flip_Flop = True AND C.Drive_Strength = 4
```

***ChipXplore* Answer (49 Seconds)**

Here's a breakdown for each parameter across libraries:
1. Leakage Power (in nW)
- sky130_fd_sc_hd: Ranges from 0.008783073 to 0.01396726. ....
2. Clock Pin Capacitance (in pF)
- sky130_fd_sc_hd: Ranges from 0.001774 to 0.002439. .....
4. Average Propagation Delay (in ns)
- sky130_fd_sc_hd: Ranges from 0.552643 to 0.754195.
- sky130_fd_sc_hs: Ranges from 0.273567 to 0.303940 ...
### Summary
-Average Propagation Delay: The 'sky130_fd_sc_hs' shows the lowest delays, indicating potentially faster performance, whereas 'sky130_fd_sc_hd' shows higher delays, which might be a trade-off for lower power consumption or smaller area. .....

# Use Cases: Analyzing Nets

**(b) User Question**

List nets exceeding thresholds: routed length > 190 microns or fanout > 20, excluding power and ground nets. Analyze the nets' name, fanout, routed length, and metal layers.

**Generated Cypher**

```
MATCH (d:Design {stage: 'routing'})-[:CONTAINS_NET]->(n:Net) WHERE
n.signal_type <> 'POWER' AND n.signal_type <> 'GROUND' AND
(n.routed_length > 190 OR n.fanout > 20) RETURN n.net_name, n.fanout,
n.routed_length, [s IN [(n)-[:ROUTED_ON]->(seg:Segment) | seg.layer] | s]
```

***ChipXplore* Answer (24 Seconds)**

Here is a breakdown of the analyzed data for each net:
1. Net Name: clknet_2_3__leaf_clk_i
- Fanout: 8.0
- Routed Length: 246.54 microns
- Metal Layers: ['met3', 'met1', 'met2'] ....
4. Net Name: _1029_
- Fanout: 6.0
- Routed Length: 192.75 microns
- Metal Layers: ['met3', 'met2', 'met1']

# Use Cases: Area and Power Tradeoffs

**(c) User Question**

What would be the total increase in area and static power if we replaced buffer cells connected to the 'rst_i' net in the current design with the buf_2 cell from the high speed library in the PDK ?

**Generated SQL**

```
SELECT Area, Cell_Leakage_Power AS Static_Power FROM Cells WHERE Name
= 'sky130_fd_sc_hs__buf_2' AND Condition_ID = ( SELECT Condition_ID
FROM Operating_Conditions WHERE Temperature = 25.0 AND Voltage = 1.8
AND Cell_Library = 'sky130_fd_sc_hs' );
```

**Generated Cypher**

```
MATCH (d:Design {stage: 'routing'})-[:CONTAINS_NET]->(n:Net {net_name
:'rst_i'})-[:CONNECTED_TO]->(c:Cell) WHERE c.is_buf = true RETURN c
```

***ChipXplore* Answer (52 Seconds)**

The existing buffer cell has an area of 3.7536 and a static power of 7.016e-5. The 'sky130_fd_sc_hs__buf_2' cell has an area of 7.992 and a static power of 0.07495. The increase in area per cell is 7.992 - 3.7536 = 4.2384, and the increase in static power per cell is 0.07495 - 7.016e-5. The total increase in area would be 4.2384 and the total increase in static power would be approximately 0.07495.

# Experimental Results: Evaluation Set

→ **Evaluation Set:**

**117** questions and their corresponding ground truth query.

→ **Accuracy Metrics:**

1) **Execution Accuracy (EX):**

LLM Generated Query returns same result as the ground truth query .

$$EX = \frac{\sum_{i=1}^{N} \mathbb{1}(V_i, \hat{V}_i)}{N}, \quad \mathbb{1}(V_i, \hat{V}_i) = \begin{cases} 1, & \text{if } V_i = \hat{V}_i \\ 0, & \text{if } V_i \neq \hat{V}_i \end{cases}$$

$\hat{V}_i$ Result returned from executing LLM generated query.

$V_i$ Result returned from ground truth query.

Table I: Evaluation set statistics, showing the number of queries and clause occurrences.

| Statistic | PDK | | | Design[1] (DEF) | Cross-Database* | Total |
|---|---|---|---|---|---|---|
| | TechLEF | LEF | Liberty | | | |
| Total Questions | 23 | 23 | 28 | 35 | 8 | 117 |
| **Clause Occurrences in Ground Truth Queries:** | | | | | | |
| JOIN | 3 | 10 | 7 | - | 5 | 25 |
| ORDER BY | 2 | 7 | 8 | 11 | 6 | 34 |
| WHERE | 26 | 40 | 79 | 11 | 10 | 166 |
| GROUP BY | 6 | 5 | 4 | - | 4 | 19 |
| Aggregation Functions | 17 | 14 | 26 | 18 | 8 | 83 |
| Sub-queries | 3 | 17 | 41 | 4 | 7 | 72 |
| Avg. Query Length (chars) | 161 | 189 | 327 | 129 | 256 | 215 |
| Max. Query Length (chars) | 452 | 661 | 761 | 258 | 893 | 975 |

[1] Design uses Cypher queries, while PDK uses SQL.
* Cross-Database questions require querying both the PDK and design database.

# Main Results

(EX): Execution Accuracy

(VES): Valid Efficiency Score

No Agents

No Planner & No Dispatcher & No Selector

No Planner & No Dispatcher

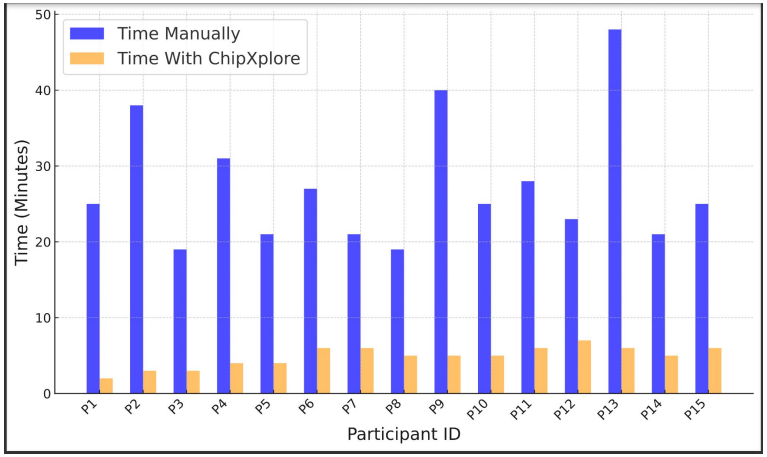
| Workflow | Model | Open-Source | PDK (SQL) | | Design (Cypher) | | Cross-Database | | Overall | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | EX (%) | VES (%) | EX (%) | VES (%) | EX (%) | VES (%) | EX (%) | VES (%) |
| Vanilla-RAG | GPT-4 | × | | | | | | | | |
| | LLAMA-3.3-70B | ✓ | | | | | | | | |
| | DEEPSEEK-V3 | ✓ | | | | | | | | |
| DIN-SQL [17] | GPT-4 | × | | | | | | | | |
| | LLAMA-3.3-70B | ✓ | | | | | | | | |
| | DEEPSEEK-V3 | ✓ | | | | | | | | |
| MAC-SQL [18] | GPT-4 | × | | | | | | | | |
| | LLAMA-3.3-70B | ✓ | | | | | | | | |
| | DEEPSEEK-V3 | ✓ | | | | | | | | |
| **ChipXplore (Ours)** | GPT-4 | × | | | | | | | | |
| | LLAMA-3.3-70B | ✓ | | | | | | | | |
| | DEEPSEEK-V3 | ✓ | | | | | | | | |
| | **LLAMA-3.3-CHIPXPLORE-70B*** | ✓ | | | | | | | | |

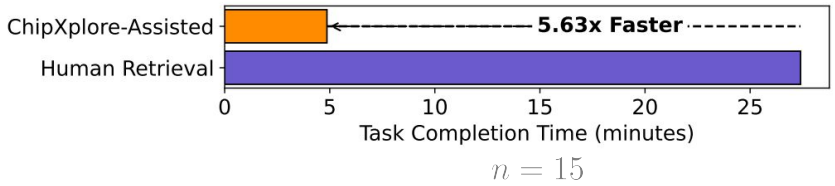
* The `LlaMa-3.3-ChipXplore-70b` model is finetuned on our SQL schema for PDK queries and generates SQL queries, while `LlaMa-3.3-70b` handles routing, schema selection, and cypher query generation.

➔ ***ChipXplore + DeepSeek-V3*** achieved the highest overall Execution Accuracy **97.39%**

# User Study: Productivity Impact

- User study with **15** Electrical and Computer Engineering graduate students.
- Participants completed **4** retrieval tasks; once manually and once using the framework




➔ ***ChipXplore*** enhanced task completion time by **5.63x ,** and reduced errors by **5.25x**

# Thank You!

https://github.com/scale-lab/ChipXplore

Find Us

School of Engineering
BROWN UNIVERSITY

# Scalability Concerns

- **In-memory** storage scales better than Disk-storage