RK3588 TTS

#whoami - Martin

- 本職： C++ / 高速計算 / 系統軟體
- 開源開發者
- 做別人不做的事

- 維護 C++ 網頁框架 - drogonframework/drogon
- 為開源軟體加上 OpenBSD 支援
- 時不時做 AI 相關研究

GitHub: https://github.com/marty1885

Blog: https://clehaxze.tw/

Disclaimer

- The opinions expressed in these slides are personal and do not represent the views or opinions of my employer. Any references to specific products, services, or organizations are for illustrative purposes only and do not imply endorsement. Please consult with appropriate professionals or your organization for specific advice related to your circumstances.

RK3588

- 中國 Rockchip 公司旗艦晶片
- 8 核心 (4 大 + 4 小)
- 外加 3 核心 NPU (神經網路處理器)
- 第三方 Ubuntu 22.04 移植
- 低功耗，最高 7W

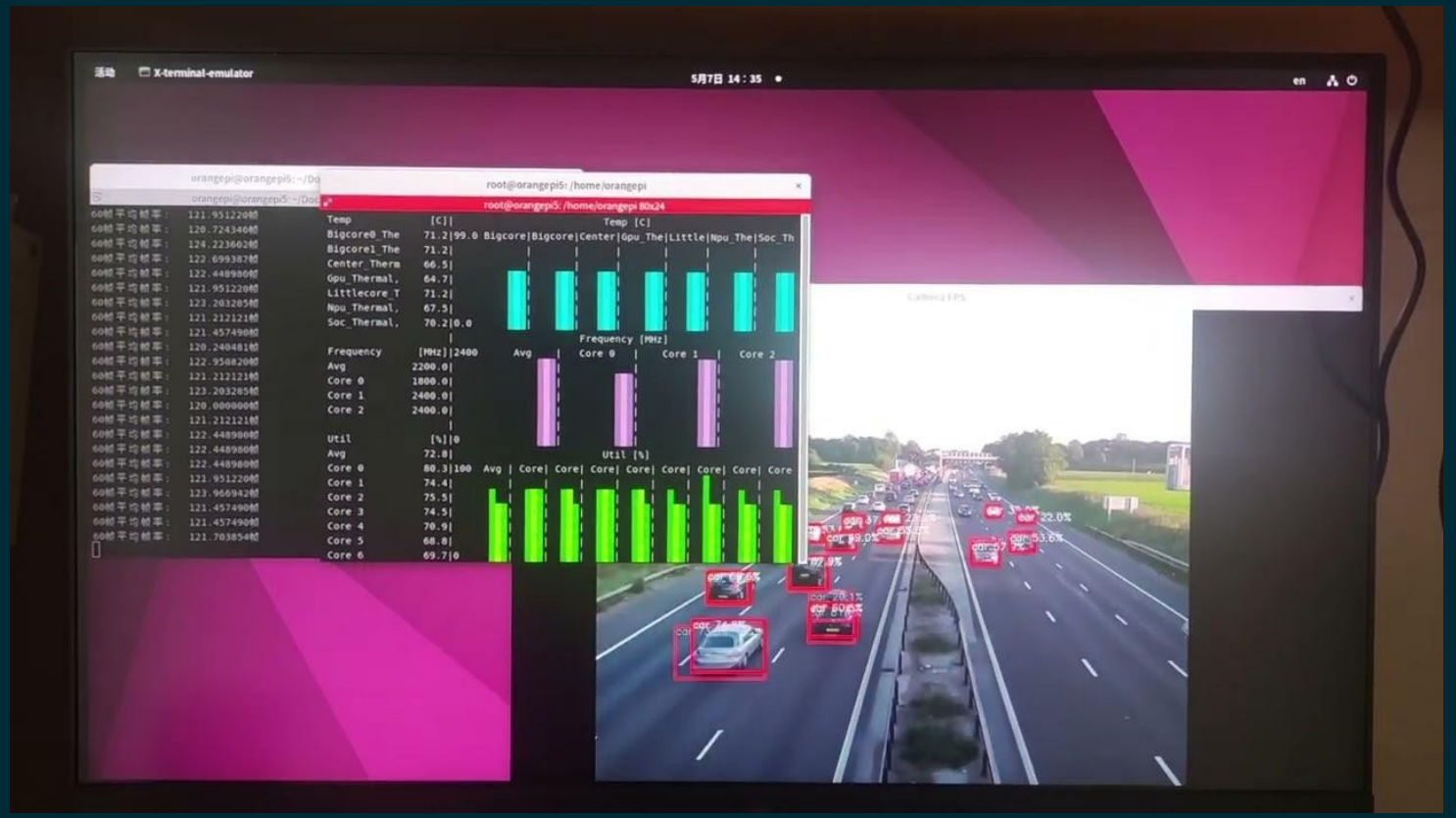
- 封閉 SDK :(

- 買來實驗跑語言模型。但那是另一個故事

RK3588 (cont..)

- NPU 設計跑 YOLO 或是 ResNet 等視覺模型
- 語音合成在設計功能之外

Credit: 李安 under fair use. YouTube VID: wwRSP9ucbhw



Motivation

- 目標：自建數位助理 - 的其中一個元件
- 問題：顯卡太熱了，無法忍受夏天擺 300W 熱源在房間

Piper

- https://github.com/rhasspy/piper
- 高品質，快速的語音合成
- CPU 上約 5.5x 合成速度
- GPU 上約 40~100x 合成速度

- RK3588 CPU: 1.1x 合成速度

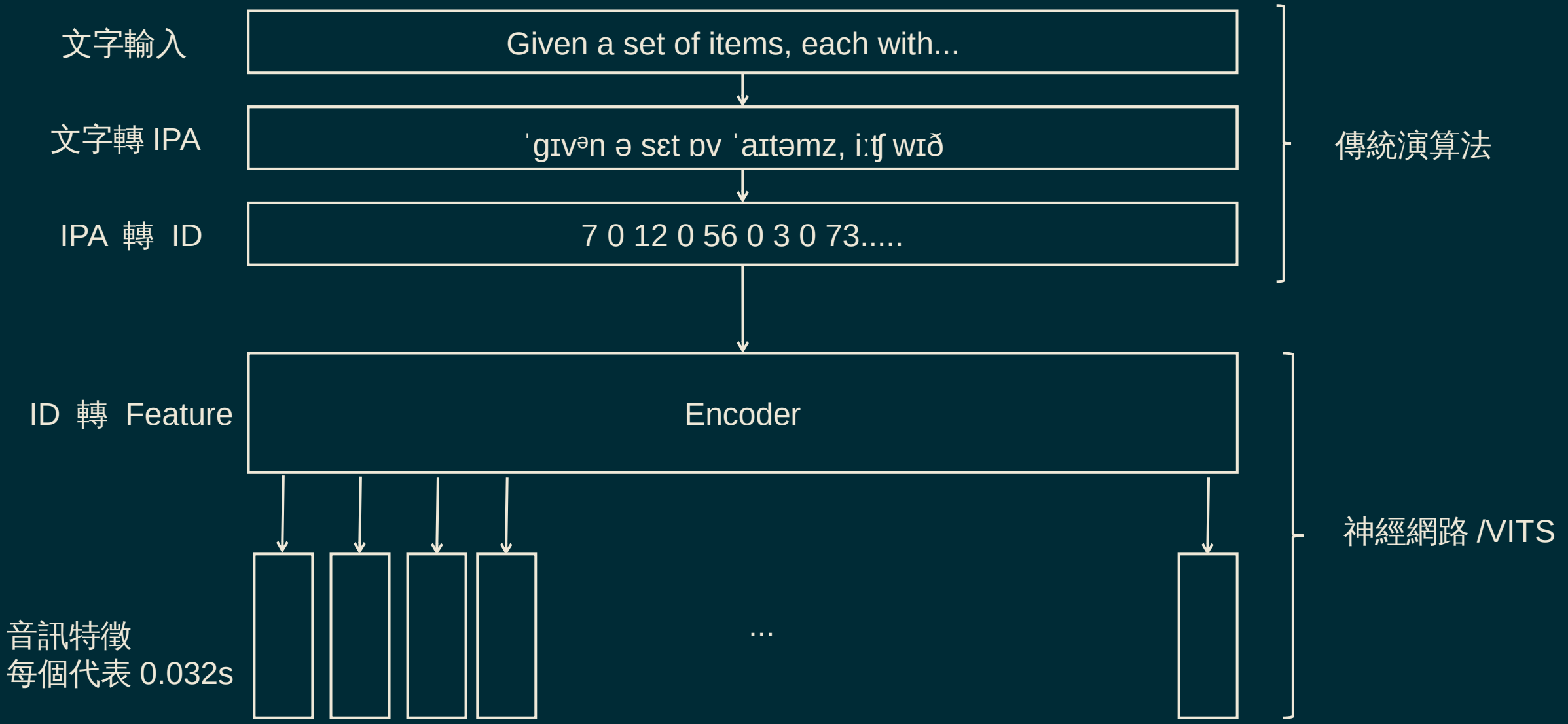
- 更嚴重的問題： Piper 以句子為單位合成。一句 1s 則要等 0.9 秒才有輸出
- Horrible UX

Can we do better?

- “ONNX streaming support” - piper #255
- 把 Piper 拆分，分割成兩個部份
- Encoder & decoder
- Encoder 還是沒辦法跑在 NPU 上
- 但 Decoder 可以
- 然後 Decoder 佔了大多數時間 !!
- 還把合成變成以 0.032s 為單位 !!

- 解決了延遲問題

Piper streaming archicture

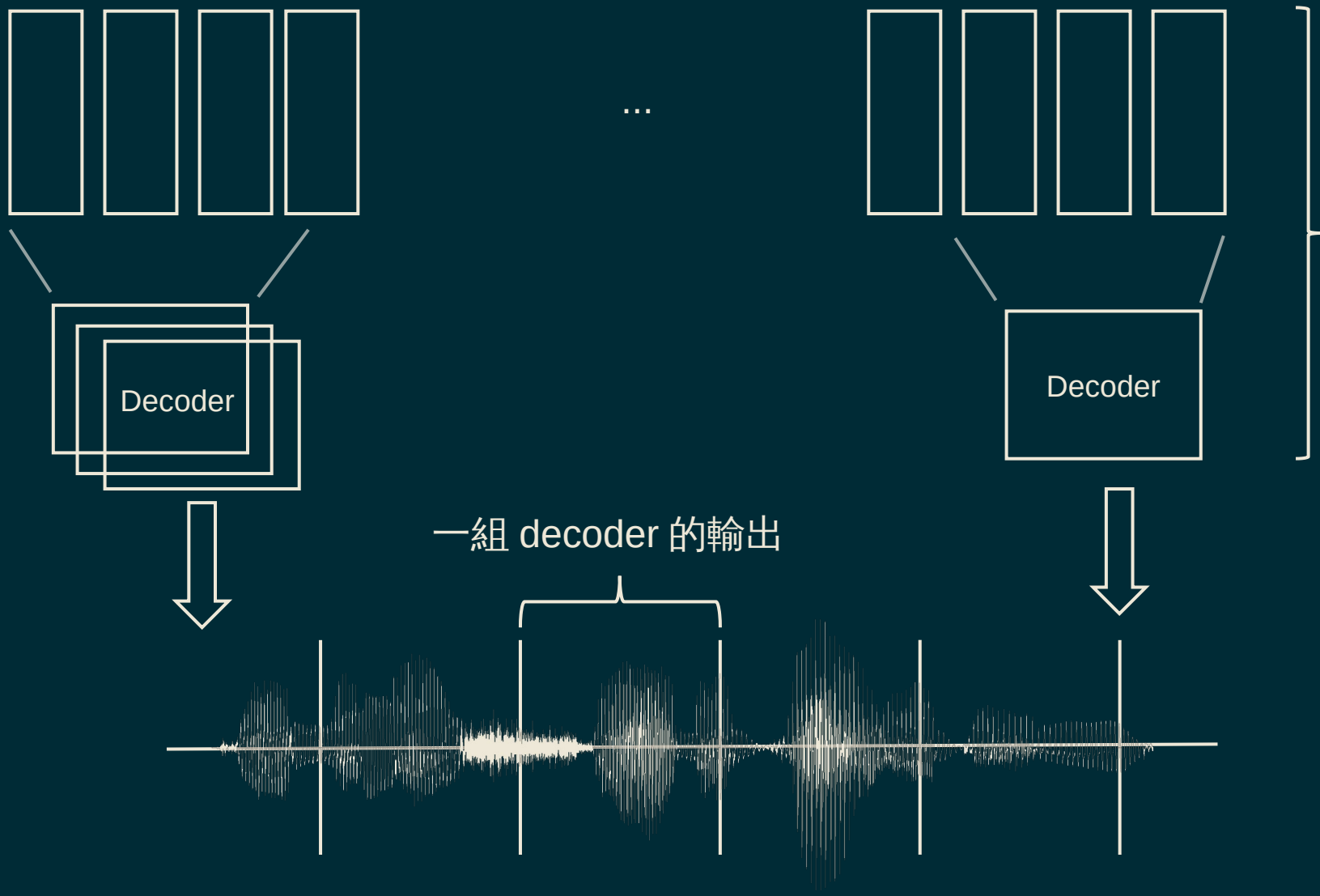


Piper streaming archicture (cont.)

音訊特徵
每個代表 0.032s

以每 N 個特徵為一組
Decoder 合成語音

組合成為連續的語音



We can do better

- 有效加速的基本要求
  - 可以加速的東西慢
  - 慢的東西佔總時間很長
- decoder 結構很像視覺模型
  - 輸入一張 2D 矩陣（長 x 寬 / 特徵 x 特徵維度）
  - 輸出一串數字（是不同物品的機率 / 音訊）
  - 裡面都是捲積 (Convolution)

- Decoder 慢
- Decoder 長得很像視覺模型

If it walks like a duck and it quacks like a duck

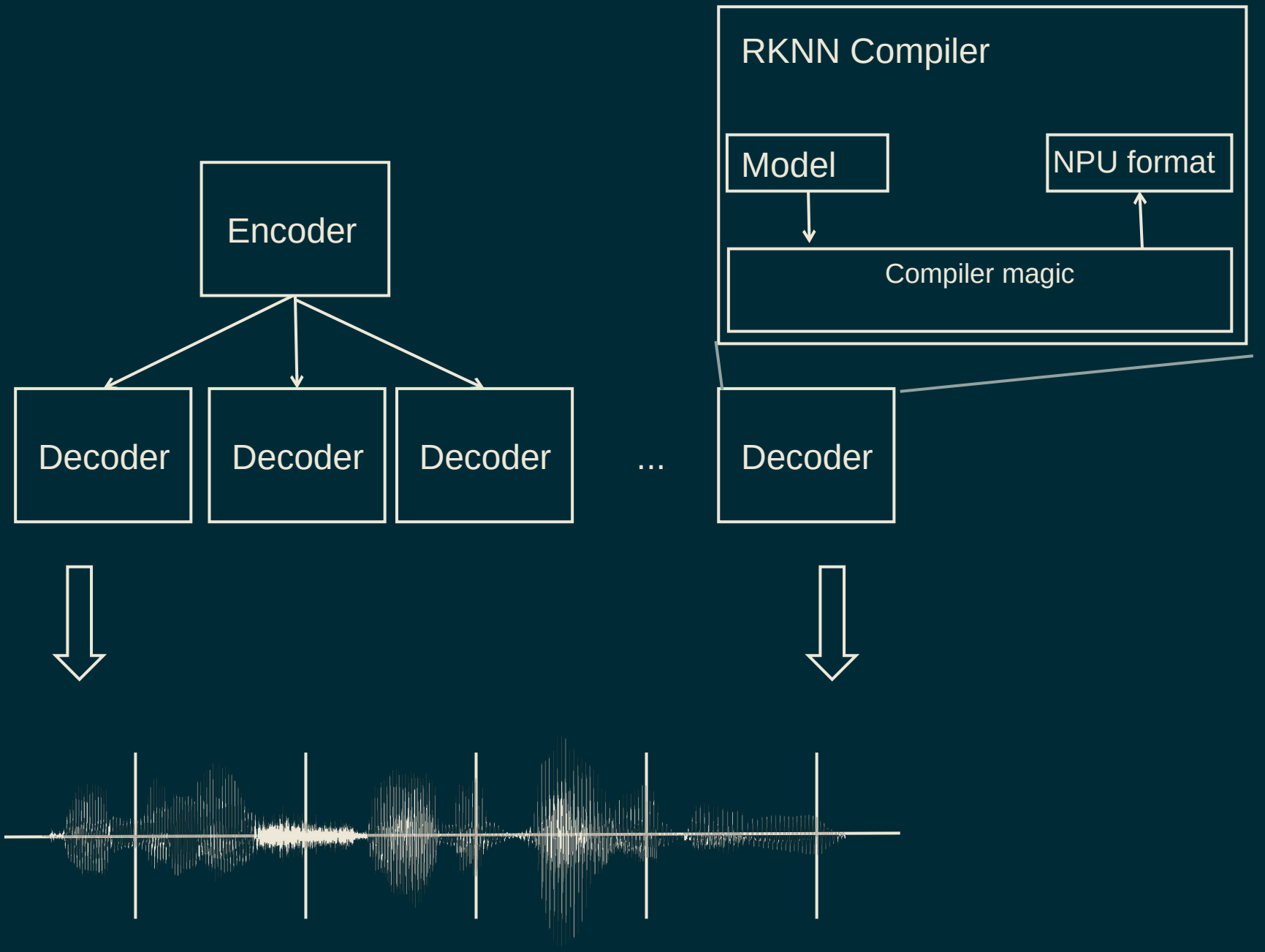
- 那它一定是隻鴨子

- 結構夠像就有機會動
- Why not try?

Much hacking later

差太遠，
繼續跑在 CPU 上

他們夠像，交
給 NPU



拿回一樣的音訊

Done

- https://github.com/marty1885/paroli
- NPU 上為 9x 合成速度
- 比桌機 CPU 還快 !!!!

- 技術可以推進 AI 應用的 UX
- 降低功耗、可應用在本來不可能的地方（家用等等）
- 因為開源與社群貢獻才做的到

Future work

- AMD 今天釋出了 XDNA driver for Linux 6.7
- Intel 14th Gen 也有 NPU

Thank you

- More to come