GUESS & SKETCH: Language Model Guided Transpilation

Celine Lee[1*], Abdulrahman Mahmoud[2*], Michal Kurek[2],
Simone Campanoni[3], David Brooks[2*], Stephen Chong[2*], Gu-Yeon Wei[2*], Alexander M. Rush[1*],
Cornell University[1], Harvard University[2], Northwestern University[3]

Motivation

Translate assembly code with language models and program synthesis.

Assembly code is composed of computer hardware-specific operations divided into basic blocks.

```
        .arch armv8-a
        .file   "problem1.c"
        .text
        .section        .rodata
        .align  3
.LC0:
        .string "%d\n"
        .text
        .align  2
        .global main
        .type   main, %function
main:
.LFB0:
        .cfi_startproc
        stp     x29, x30, [sp,
-32]!
        .cfi_def_cfa_offset 32
        .cfi_offset 29, -32
        .cfi_offset 30, -24
        mov     x29, sp
        str     wzr, [sp, 16]
        str     wzr, [sp, 20]
        str     wzr, [sp, 24]
        str     wzr, [sp, 28]
        b       .L2
.L6:
        ldr     w2, [sp, 28]
        mov     w0, 21846
        movk    w0, 0x5555, lsl
16
        smull   x0, w2, w0
        lsr     x1, x0, 32
                ...
```

Hand-engineered transpilers are expensive to make.

LMs produce plausible outputs, but do not guarantee correctness.

Symbolic methods guarantee correctness but scale poorly.

GUESS & SKETCH takes a **neurosymbolic approach** that leverages the scalability of language models with the correctness of symbolic solvers.

Objective

$$x \equiv y : \forall d \in \mathcal{D} : P_x(d) = P_y(d)$$

For input program P_x represented as sequence x, produce the semantically-equivalent P_y represented as sequence y.

Semantic equivalence is measured by execution equivalence on the domain of all program inputs $\mathcal{D}$.

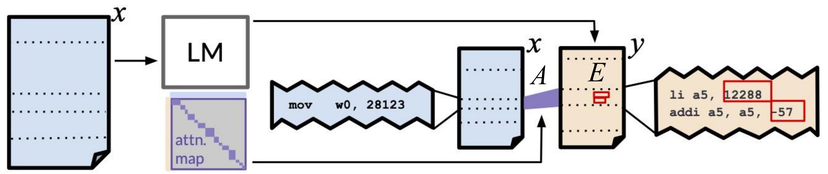
GUESS: Propose translations, detect potential errors

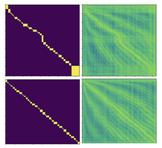
$$x \rightarrow (y, A, E)$$

For input sequence x, produce tuples:

(guessed translation y, subseq.s alignment A, potential token-level errors E)

$$\arg\max_{y \in \mathcal{V}^L} \prod_t p(y_t | y_{<t}, x) \qquad A \in \mathcal{B}_x^{|\mathcal{B}_y|} \qquad E \in \{0,1\}^{|y|}$$





Alignment A is extracted between subsequences $b \in \mathcal{B}$ in x and y by the highest aggregate attention score.

Potential errors E are detected by probability $p(y_j | y_{<j}, x) < \gamma$ and generating out-of-scope references.

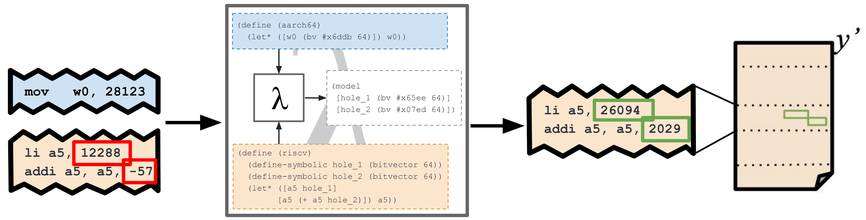
SKETCH: Correct errors

$$(y, A, E) \rightarrow y'$$

Sketches are created from each potentially-erroneous subsequence y_b by replacing each position $j \in b$ that also satisfies $E_j \neq 0$ with a hole $\bullet$.

Solve sketches by finding the mapping ϕ that populates all holes of the sketch to satisfy the correctness spec set by x_{b_x} where $A_b = b_x$:

$$\forall d \in \mathcal{D} : P_{x_{b_x}}(d) = P_{\phi(s)}(d)$$



Experiments

Train and test sets for transpilation are compiled to the ARMv8 and RISC-V architectures under the `-O0` optimization flag.

	#	Avg. # lines	In	Out
Unix Commands	11	96	✓	✓
Project Euler	45	159		✓
Benchmarks Game	16	484	✓	✓

Results

GUESS & SKETCH transpiles 57.6% more examples than GPT-4 and 39.6% more examples than an engineered transpiler.

	RISC-V to ARMv8			ARMv8 to RISC-V		
	Proj. Euler	Benchmx	Unix Cmds	Proj. Euler	Benchmx	Unix Cmds
Few-shot (GPT4)	11.1%	0	18.2%	4.44%	0	27.3%
Transpiler	-	-	-	24.4%	12.5%	54.5%
FT StarCoder	8.9%	0	36.4%	8.9%	0	36.4%
FT CodeLlama	11.1%	0	36.4%	2.2%	0	36.4%
Enc-Decoder	68.9%	6.3%	36.4%	66.7%	6.25%	**81.2%**
GUESS & SKETCH	**80.0%**	**18.8%**	**81.2%**	**75.6%**	**25.0%**	**81.2%**

With the symbolic component, fewer samples are needed from the language model.

	Project Euler	
	RISC-V to ARMv8	ARMv8 to RISC-V
Encoder-Decoder	30.1	34.3
GUESS & SKETCH	**21.3**	**25.3**

Error Analysis

- Pre-trained LLMs produce instructions from different ISAs.
- Many assembly files exceed LLM context window.
- Attention alignment is not 100% accurate.
- Resulting search space for symbolic component can be too large to resolve in a reasonable time frame.

C. Lee, A. Mahmoud, M. Kurek, S. Campanoni, D. Brooks, S. Chong, G-Y Wei, A. M. Rush, GUESS & SKETCH: Language Model Guided Transpilation. In *The Twelfth International Conference on Learning Representations*, 2024.