Fixslicing AES-like Ciphers

New bitsliced AES speed records on ARM-Cortex M and RISC-V

Alexandre Adomnicai and Thomas Peyrin

Nanyang Technological University, Singapore
Temasek Laboratories, Singapore
firstname.lastname@ntu.edu.sg

Abstract. The fixslicing implementation strategy was originally introduced as a new representation for the hardware-oriented GIFT block cipher to achieve very efficient software constant-time implementations. In this article, we show that the fundamental idea underlying the fixslicing technique is not of interest only for GIFT, but can be applied to other ciphers as well. Especially, we study the benefits of fixslicing in the case of AES and show that it allows to reduce by 52% the amount of operations required by the linear layer when compared to the current fastest bitsliced implementation on 32-bit platforms. Overall, we report that fixsliced AES-128 allows to reach 80 and 91 cycles per byte on ARM Cortex-M and E31 RISC-V processors respectively (assuming pre-computed round keys), improving the previous records on those platforms by 21% and 26%. In order to highlight that our work also directly improves masked implementations that rely on bitslicing, we report implementation results when integrating first-order masking that outperform by 12% the fastest results reported in the literature on ARM Cortex-M4. Finally, we demonstrate the genericity of the fixslicing technique for AES-like designs by applying it to the Skinny-128 tweakable block ciphers.

Keywords: AES · ARM · RISC-V · Implementation · Bitslicing · Fixslicing

1 Introduction

Since the selection of the Rijndael block cipher as the Advanced Encryption Standard (AES) [DR02] in 2001, optimized implementations of this algorithm attracted a lot of interest over the past two decades. If AES can be efficiently implemented using look-up tables, the table accesses being key and data-dependent lead to cache-timing attacks [Ber05, BM06]. With these vulnerabilities in mind, cryptographers came up with constant-time implementations by taking advantage of vector permute instructions [Ham09] or bitslicing [MN07, Kön08, KS09]. To meet the need for efficient and secure implementations, Intel and AMD added the set of x86 instructions AES-NI [Gue08] to implement AES using dedicated hardware circuits. However, because such dedicated instructions are not necessarily available on a given platform, the study of efficient constant-time AES implementations is still an active research topic, especially on microprocessors used in low-end embedded devices because of their limited computational resources. Although there are undergoing initiatives that intend to provide lightweight alternatives to AES for such platforms (e.g. the NIST LWC project [MBTM17]), it will probably still be widely deployed in the near future for security guarantees and compliance reasons. To date, the fastest constant-time AES implementation on 32-bit reduced instruction set computer (RISC) is is the one from Schwabe and Stoffelen [SS16] that runs at 101 cycles per byte (cpb) on ARM Cortex-M3 by processing 2 blocks in parallel. It was also ported to the 32-bit RISC-V architecture and results in 124 cpb on this platform [Sto19]. This implementation, which
relies on bitslicing, also serves as a basis for some of the current best results reported in the literature when integrating countermeasures against power/electromagnetic side-channel attacks [GSDM+19]. One notable feature of this implementation is that about 40% and 55% of the cycles are spent for the linear layer on ARM Cortex-M and RISC-V, respectively. Clever optimizations of software linear layer implementations have already been addressed for some ciphers. In [RAL17], the authors introduce an alternative representation of the PRESENT block cipher [BKL+07] over 2 rounds that allows to speed up the performance by a factor 8 on ARM Cortex-M. More recently, a similar approach named fixslicing [ANP20] has been applied to the GIFT family of block ciphers [BPP+17], enhancing the performance by a factor 7 on ARM Cortex-M when compared to naive bitslicing. Those works highlight that the performance of a bitsliced implementation not only depends on the way the bits are packed within registers, but also on possible alternative representations of the cipher. Because such optimizations have only been applied to Substitution-bitPermutation Network (SbPN) such as GIFT and PRESENT to date, it remains unclear if it would be of interest for other designs. However, the generic aspect of fixslicing tends to indicate that this concept might be more widely applicable.

Our contributions. In this article, we intend to enhance the current best speed results for constant-time implementations of AES on embedded 32-bit platforms. By analyzing the performance of the current fastest implementation, we note that more of 30% of the instructions are dedicated to the ShiftRows layer. As a first step to minimize the cost of this operation, we push bitsliced implementations of AES to their limit on 32-bit platforms by introducing a new bitsliced representation that we call barrel-shiftrows. The advantage of this representation is the ability to compute the ShiftRows simply using 32-bit rotations while not impacting the MixColumns efficiency. On the other hand, it requires to process 8 blocks in parallel which might not be well suited to handle efficiently a small amount of data, as it is often the case for embedded devices. Therefore, instead of focusing on a new way to pack the bits within registers, we investigate the benefits of fixslicing in the case of AES. We show that the fundamental idea underlying this concept is not only of interest for SbPN designs, but can be applied to other ciphers as well. Indeed, fixslicing allows to reduce by 52% the amount of operations required by the linear layer when compared to the current fastest bitsliced implementation on 32-bit platforms. All in all, we report that fixsliced AES-128 reaches 80 and 91 cpb on ARM Cortex-M and E31 RISC-V processors respectively (assuming pre-computed round keys), improving the previous records on those platforms by 21% and 26%. Those results require the ability to process two blocks simultaneously and therefore apply to all parallelizable modes of operation (e.g. CTR, GCM). Our work directly improves the prior reported results for first-order masked AES on ARM Cortex-M4 by 12%, with 187 cpb. Finally, we highlight that the fixsliced approach can be applied to other AES-like designs by illustrating an application to the Skinny-128 tweakable block ciphers. All our implementations are available in the public domain at https://github.com/aadomn/aes.

2 Preliminaries

2.1 AES overview

AES is a 128-bit block cipher that can be instantiated using three different key lengths: 128, 192 or 256 bits, resulting in three corresponding versions: AES-128, AES-192 and AES-256. All versions rely on the same round function, which is applied 10, 12 and 14 times for AES-128, AES-192 and AES-256, respectively. The round function, which operates on the internal state viewed as a $4 \times 4$ matrix of elements in the finite field defined by the irreducible polynomial $x^8 + x^4 + x^3 + x + 1$ over GF(2), consists in the following four
Fixslicing AES-like Ciphers

operations:
- **SubBytes**: applies the same 8-bit S-box to each byte of the internal state
- **ShiftRows**: shifts the $i$-th row left by $i$ bytes
- **MixColumns**: multiplies each column with a diffusion matrix over $\text{GF}(2^8)$
- **AddRoundKey**: adds a 128-bit round key to the internal state.

![The AES round function.](image)

The AES round function is illustrated in Figure 1. Note that an additional **AddRoundKey** is performed at the very beginning of the first round, and that the **MixColumns** operation is omitted during the last round. The encryption key is expanded into round keys using a key schedule algorithm, whose round function is depicted in Figure 2 for each AES version. Note that a round constant is also incorporated in each round keys, we refer to [DR02] for more details.

![Key schedule round functions for each AES version.](image)

**2.2 Bitslicing the AES**

Bitslicing is a software implementation technique where the computation of a function is reduced to logic gates (e.g. **AND**, **XOR**, **OR**, **NOT**), allowing to execute as many instances in parallel as the CPU’s register width. It was originally introduced as an efficient way to implement the DES block cipher [Bih97, Kwa00] before being considered as a generic technique to achieve fast constant-time implementations. Many bitsliced AES implementations followed [RSD06, MN07, Kön08] where the fastest of them was introduced by Käser and Schwabe [KS09] allowing to reach 6.9 cpb for AES-128 on Intel Core i7 processors by processing 8 blocks in parallel as depicted in Figure 3.
While bitsliced AES implementations aroused less interest on high-end processors since the deployment of the AES-NI instruction set, it still attracts a lot of attention for platforms that do not enjoy AES hardware acceleration, such as low-end microprocessors. Although the most constrained microprocessors do not necessarily have any internal cache memory (e.g. ARM Cortex-M3), it is possible for a system on chip design to integrate a system level cache, making cache-timing attacks a threat. Moreover, because embedded platforms are typical targets for side-channel attacks such as differential power/electromagnetic analysis, relying on an implementation that works at the gate level facilitates the integration of Boolean masking as a countermeasure.

On 32-bit platforms, the most efficient bitsliced AES implementation reported in the literature is the one from Schwabe and Stoffelen [SS16] allowing to reach 101 cpb on ARM Cortex-M3. It was also ported to the 32-bit RISC-V architecture and results in 124 cpb on E31 processors [Sto19]. Their implementation heavily relies on [KS09] by adapting it to 32-bit registers instead of 128-bit ones as depicted in Figure 4.

The advantage of this representation is the ability to compute the MixColumns operation using only 27 exclusive-ORs and 16 rotations. Indeed, because each byte in the internal state is an element of GF(2)/x^8 + x^4 + x^3 + x + 1, multiplication by 2 is achieved by a left...
shift and conditional masking with (00011011)₂ whenever the most significant bit (MSB) equals 1. Since $R_0$ contains the MSB of each byte, one has simply to add it to the four corresponding registers. Moreover, because the bitsliced representation of the internal state is row-wise, adding an adjacent element in the column simply corresponds to an exclusive-OR combined with a rotation. Therefore, the entire MixColumns computation can be achieved in the following way:

\[
\begin{align*}
R'_0 &= (R_1 \oplus R_3^{20^8}) \oplus R_0^{20^8} \oplus (R_0 \oplus R_6^{20^8})^{20^8} \\
R'_1 &= (R_2 \oplus R_4^{20^8}) \oplus R_1^{20^8} \oplus (R_1 \oplus R_3^{20^8})^{20^8} \\
R'_2 &= (R_3 \oplus R_5^{20^8}) \oplus R_2^{20^8} \oplus (R_2 \oplus R_4^{20^8})^{20^8} \\
R'_3 &= (R_4 \oplus R_6^{20^8}) \oplus R_3^{20^8} \oplus (R_3 \oplus R_5^{20^8})^{20^8} \oplus (R_0 \oplus R_6^{20^8}) \\
R'_4 &= (R_5 \oplus R_7^{20^8}) \oplus R_4^{20^8} \oplus (R_4 \oplus R_6^{20^8})^{20^8} \oplus (R_0 \oplus R_6^{20^8}) \\
R'_5 &= (R_6 \oplus R_0^{20^8}) \oplus R_5^{20^8} \oplus (R_5 \oplus R_7^{20^8})^{20^8} \\
R'_6 &= (R_7 \oplus R_2^{20^8}) \oplus R_6^{20^8} \oplus (R_6 \oplus R_0^{20^8})^{20^8} \oplus (R_0 \oplus R_0^{20^8}) \\
R'_7 &= (R_0 \oplus R_4^{20^8}) \oplus R_7^{20^8} \oplus (R_7 \oplus R_2^{20^8})^{20^8}
\end{align*}
\]

where $R_i^{20^j}$ refers to a rotation of $R_i$ by $j$ bits to the right. Note that on ARM, thanks to the inline barrel shifter, the rotations can be computed for free resulting in only 27 1-cycle instructions in total.

While the row-wise bitsliced representation allows an efficient MixColumns implementation, it is less suited regarding the ShiftRows operation. When considering 8 blocks using 128-bit registers, the ShiftRows corresponds to a byte-level permutation on each register, which can be efficiently computed on Intel using the SSSE3 byte shuffle instruction `pshufb`. However for the 32-bit version, according to the representation depicted in Figure 4, the ShiftRows requires to compute byte-wise rotations. This can be achieved by means of 6 OR instructions, 7 AND instructions and 6 logical shifts are required per register as shown in Listing 1. Note that [SS16] uses bitfield extract instructions for their ARM implementation but it does not achieve better performance anyway.

Listing 1: C code to apply the ShiftRows on a slice $r$ according to the bitsliced representation in Figure 4.

On ARM, thanks to the inline barrel shifter, it results in $(6 + 7) \times 8 = 104$ 1-cycle instructions per ShiftRows, leading to $(104 \times 10)/32 = 32.5$ cpb which is 32% of the overall AES-128 performance reported on ARM Cortex-M. On RISC-V it corresponds to $19 \times 8 = 152$ 1-cycle instructions per ShiftRows, leading to $152 \times 10/32 = 47.5$ cpb which is 38% of the overall AES-128 performance reported on E31 RISC-V processors. However, note that this is not optimal: after having uploaded a preliminary version of our work online, Dettman highlighted that it can be done more efficiently\(^1\) as detailed in Listing 2. Because the implementations have not been patched yet at the time of writing, we do not

---

\(^{1}\) Improved aes128ctrbs shift row suggestion
consider this optimization for our benchmarks since there is no practical results available.
Instead we briefly discuss some estimates in Section 5.3.

\begin{lstlisting}[language=C]
SWAPMOVE(r, r, 0x030f0c00, 4);
SWAPMOVE(r, r, 0x33003300, 2);
\end{lstlisting}

\section*{3 A new ShiftRows-friendly representation}

A straightforward way to reduce the cost of the \texttt{ShiftRows} operation is to keep a row-wise bitsliced representation and to isolate each row in distinct registers, so that byte-wise rotations are replaced by word-wise rotations. However on 32-bit platforms, it requires 32 registers to store the internal state by processing 8 blocks in parallel as illustrated in Figure 5. We refer to this representation as \textit{barrel-shiftrows} since it allows to compute the \texttt{ShiftRows} using only 24 32-bit rotation. On ARM, it means that the \texttt{ShiftRows} can be actually computed for free by using the inline barrel shifter. However as there are only 14 general-purpose registers available, one would have to deal with numerous memory accesses throughout the AES processing. At first glance, it is not clear how it would perform when compared to [SS16]. On the other hand, the barrel-shiftrows representation could be more valuable on platforms that embed more registers and that do not come with any rotation instruction (e.g. RV32I). Indeed, the \texttt{MixColumns} no longer requires rotations but only exclusive-ORs since the different bytes within a column are now stored in distinct registers. Therefore, instead of computing a rotation to ensure that all bytes within the column are properly aligned, one has just to perform an exclusive-OR with the corresponding registers as detailed in Equation 2:

\begin{equation}
\begin{align*}
R'_i &= R_{i+1} \oplus R_{i+3} \oplus R_{i+8} \oplus R_{i+16} \oplus R_{i+24} \\
R'_{i+1} &= R_{i+2} \oplus R_{i+9} \oplus R_{i+17} \oplus R_{i+25} \\
R'_{i+2} &= R_{i+3} \oplus R_{i+11} \oplus R_{i+18} \oplus R_{i+26} \\
R'_{i+3} &= R_{i+4} \oplus R_{i+12} \oplus R_{i+19} \oplus R_{i+27} \oplus (R_i \oplus R_{i+8}) \\
R'_{i+4} &= R_{i+5} \oplus R_{i+13} \oplus R_{i+20} \oplus R_{i+28} \oplus (R_i \oplus R_{i+8}) \\
R'_{i+5} &= R_{i+6} \oplus R_{i+14} \oplus R_{i+21} \oplus R_{i+29} \\
R'_{i+6} &= R_{i+7} \oplus R_{i+15} \oplus R_{i+22} \oplus R_{i+30} \oplus (R_i \oplus R_{i+8}) \\
R'_{i+7} &= (R_i \oplus R_{i+8}) \oplus R_{i+15} \oplus R_{i+23} \oplus R_{i+31}
\end{align*}
\end{equation}

for $i \in \{0, 8, 16, 24\}$ and where all subscripts are to be considered modulo 32.

Using the barrel-shiftrows representation, the \texttt{MixColumns} requires $27 \times 4 = 108$ exclusive-ORs by processing 8 blocks in parallel, while the bitsliced representation requires $16 \times 4 = 64$ additional rotations. While this is not of particular interest on ARM, this is beneficial to platforms without rotate instruction. On 32-bit platforms, the barrel-shiftrows representation might be the most efficient way to compute the \texttt{ShiftRows} operation. However it requires to process 8 blocks in parallel which can be inappropriate for communication protocols used in embedded systems that are designed to transmit small amount of data. In the next section, we look at optimizing the representation that processes only 2 blocks at a time.
Figure 5: Barrel-shiftrows representation using 32 32-bit registers $R_0, \cdots, R_{31}$ to process 8 blocks $b^0, \cdots, b^7$ in parallel where $b^i_j$ refers to the $i$-th bit of the $j$-th block.

4 Fixslicing the AES

Instead of looking for a new way to pack the bits within the registers, another interesting and promising approach is to investigate whether it would be advantageous to not follow the classical cipher representation for a few rounds. By following this strategy, it was possible to greatly enhance the performances of the GIFT block cipher in software [ANP20]. To put it in a nutshell, the authors proposed an alternative representation of the cipher over several rounds to minimize the cost of the linear layer. They call their implementation technique fixslicing as it mainly consists in fixing the bits within a register (or slice) to never move and to adjust the other slices accordingly so that the proper bits are involved in the SubBytes operation. At first glance, it seems that the fixslicing technique as originally specified is only of interest for SbPN designs which have the special property that each bit located in a slice remains in this same slice through the permutation. However, the main idea underlying the fixslicing technique, which is to rely on an alternative representation of the cipher for a few rounds while ensuring that the bits are correctly aligned for the SubBytes computation, is actually generic and might be of interest for numerous designs. In this section, we study the relevance of fixslicing with regards to the AES on 32-bit platforms.

4.1 Application to the round function

In the case of SbPN ciphers where the permutation layer simply consists of a bit permutation, the only requirements when considering an alternative representation of the cipher over several rounds are to adapt the round keys accordingly and to ensure that the bits are correctly aligned for the non-linear layer. However, for AES-like ciphers the permutation layer comprises two linear operations, namely ShiftRows as a byte permutation and MixColumns as a matrix multiplication. Therefore, it is not sufficient to just ensure that
the bits are properly aligned with regards to the SubBytes operations, it has to be done for the exclusive-ORs in the MixColumns as well. According to the bitsliced representation detailed in Figure 4, fixing one of the slices (or registers) to never move means to simply omit the ShiftRows operation throughout the entire algorithm execution. Note that to have the bits correctly aligned to perform the SubBytes in a bitsliced manner, all slices have to remain fixed. Therefore, the main issue raised by the omission of the ShiftRows permutation is to adapt the MixColumns accordingly.

Before entering the MixColumns during the first round, it is trivial that $F = SR^{-1}(S)$ where $F$, $S$ refer to the internal state in the fixsliced and classical representations respectively, and $SR$ refers to the ShiftRows permutation. Thus, to ensure the correctness of the MixColumns operation, one has to compute the ShiftRows (i.e. the corresponding byte-wise rotations) on some temporary registers, so that the proper bits are exclusive-ORed together. The calculations are detailed in Figure 6.

$$R_0 = \left( R_1 \oplus (R_1^{2^{8}} \gg 6) \right) \oplus (R_0^{2^{8}} \gg 6) \oplus (R_0^{2^{16}} \gg 4) \oplus (R_0^{2^{24}} \gg 2)$$
$$R_1 = \left( R_2 \oplus (R_2^{2^{8}} \gg 6) \right) \oplus (R_1^{2^{8}} \gg 6) \oplus (R_1^{2^{16}} \gg 4) \oplus (R_1^{2^{24}} \gg 2)$$
$$R_2 = \left( R_3 \oplus (R_3^{2^{8}} \gg 6) \right) \oplus (R_2^{2^{8}} \gg 6) \oplus (R_2^{2^{16}} \gg 4) \oplus (R_2^{2^{24}} \gg 2)$$
$$R_3 = \left( R_4 \oplus (R_4^{2^{8}} \gg 6) \right) \oplus (R_3^{2^{8}} \gg 6) \oplus (R_3^{2^{16}} \gg 4) \oplus (R_3^{2^{24}} \gg 2) \oplus (R_0 \oplus (R_0^{2^{8}} \gg 6))$$
$$R_4 = \left( R_5 \oplus (R_5^{2^{8}} \gg 6) \right) \oplus (R_4^{2^{8}} \gg 6) \oplus (R_4^{2^{16}} \gg 4) \oplus (R_4^{2^{24}} \gg 2) \oplus (R_0 \oplus (R_0^{2^{8}} \gg 6))$$
$$R_5 = \left( R_6 \oplus (R_6^{2^{8}} \gg 6) \right) \oplus (R_5^{2^{8}} \gg 6) \oplus (R_5^{2^{16}} \gg 4) \oplus (R_5^{2^{24}} \gg 2)$$
$$R_6 = \left( R_7 \oplus (R_7^{2^{8}} \gg 6) \right) \oplus (R_6^{2^{8}} \gg 6) \oplus (R_6^{2^{16}} \gg 4) \oplus (R_6^{2^{24}} \gg 2) \oplus (R_0 \oplus (R_0^{2^{8}} \gg 6))$$
$$R_7 = \left( R_0 \oplus (R_0^{2^{8}} \gg 6) \right) \oplus (R_7^{2^{8}} \gg 6) \oplus (R_7^{2^{16}} \gg 4) \oplus (R_7^{2^{24}} \gg 2)$$

**Figure 6:** Equations to compute the MixColumns during the first fixsliced round where $R_i \gg j$ refers to a byte-wise rotation of $j$ bits to the right, for all bytes within $R_i$.

Since $(R_i^{2^{16}} \gg 4) \oplus (R_i^{2^{24}} \gg 2) = (R_i \oplus (R_i^{2^{8}} \gg 6))^{2^{16}} \gg 4$, the fixsliced MixColumns detailed in Figure 6 can be computed using 27 exclusive-ORs, 16 word-wise rotations and 16 byte-wise rotations. All in all, it corresponds to 27 XOR, 32 AND and 16 OR instructions on top of 16 circular and 32 logical shifts. When compared to the classical bitsliced representation it saves 72 instructions for the entire linear layer, namely 32 OR, 24 AND and 16 logical shifts. As detailed in the rest of this section, the benefits of the fixslicing implementation strategy are even more significant during the next rounds.

Before entering the MixColumns during the second round, we now have $F = SR^{-2}(S)$ which implies that the first and third rows are aligned with the classical representation, whereas the second and fourth ones are shifted by two bytes. This is especially beneficial to the fixsliced representation as it means that just a single byte-wise rotation per register is needed as described in Figure 7. Indeed, during the first round, each row in the fixsliced internal state is delayed by one byte shift to the left in comparison to its adjacent rows. In other words, one has to shift by one position to the left the row $i$ to be aligned with the row $i + 1 \mod 4$. However, the row $i$ has to be shifted by 2 (resp. 3) positions to the left to match the row $i + 2 \mod 4$ (resp. $i + 3 \mod 4$) alignment. This is why 3 byte-wise rotations with 3 different rotation values (i.e. 6, 4 and 2) are required for each register in Figure 6. During the second round, because each row is either aligned or shifted by 2

---

2Improved AES fixslice MixColumns algorithm(s)
positions compared to all other rows, only a single byte-wise rotation by 4 bits is required per register. Therefore, the fixsliced MixColumns in the second round requires 27 XOR, 16 AND and 8 OR instructions on top of 16 circular and 16 logical shifts.

\[
R'_0 = (R_1 \oplus (R_1^{208} \ggg s 4)) \oplus (R_0^{208} \ggg s 4) \oplus (R_0 \oplus (R_0^{208} \ggg s 4))^{30-16} \\
R'_1 = (R_2 \oplus (R_2^{208} \ggg s 4)) \oplus (R_1^{208} \ggg s 4) \oplus (R_1 \oplus (R_1^{208} \ggg s 4))^{30-16} \\
R'_2 = (R_3 \oplus (R_3^{208} \ggg s 4)) \oplus (R_2^{208} \ggg s 4) \oplus (R_2 \oplus (R_2^{208} \ggg s 4))^{30-16} \\
R'_3 = (R_4 \oplus (R_4^{208} \ggg s 4)) \oplus (R_3^{208} \ggg s 4) \oplus (R_3 \oplus (R_3^{208} \ggg s 4))^{30-16} \oplus (R_0 \oplus (R_0^{208} \ggg s 4)) \\
R'_4 = (R_5 \oplus (R_5^{208} \ggg s 4)) \oplus (R_4^{208} \ggg s 4) \oplus (R_4 \oplus (R_4^{208} \ggg s 4))^{30-16} \oplus (R_0 \oplus (R_0^{208} \ggg s 4)) \\
R'_5 = (R_6 \oplus (R_6^{208} \ggg s 4)) \oplus (R_5^{208} \ggg s 4) \oplus (R_5 \oplus (R_5^{208} \ggg s 4))^{30-16} \\
R'_6 = (R_7 \oplus (R_7^{208} \ggg s 4)) \oplus (R_6^{208} \ggg s 4) \oplus (R_6 \oplus (R_6^{208} \ggg s 4))^{30-16} \oplus (R_0 \oplus (R_0^{208} \ggg s 4)) \\
R'_7 = (R_0 \oplus (R_0^{208} \ggg s 4)) \oplus (R_7^{208} \ggg s 4) \oplus (R_7 \oplus (R_7^{208} \ggg s 4))^{30-16}
\]

**Figure 7:** Equations to compute the MixColumns during the second fixsliced round where \(R_i \ggg j \) refers to a byte-wise rotation of \(j\) bits to the right, for all bytes within \(R_i\).

Due to the ShiftRows transformation, the third round configuration will be similar to the first one except that each row will be delayed by one byte shift to the right (instead of left) in comparison to its adjacent rows. Therefore the computation of the fixsliced MixColumns in the third round is the same as in the first round, with a slight modification: byte-wise rotation values have to be reversed. For instance, the update of \(R_0\) is defined by:

\[
R'_0 = (R_1 \oplus (R_1^{208} \ggg s 2)) \oplus (R_0^{208} \ggg s 2) \oplus (R_0^{20-16} \ggg s 4) \oplus (R_0^{20-24} \ggg s 6). \quad (3)
\]

Therefore, the third round requires exactly the same number of operations as the first one. In the fourth round, the fixsliced representation will be finally synchronized with the classical one for the MixColumns since \(SR^2 = Id\). As a result, one can simply compute the permutation layer using 27 XOR instructions and 16 circular shifts as detailed in Figure 1.

Consequently, our fixsliced AES description relies on a quadruple round routine where each round only differs by its implementation of the linear layer. Since only one AES version has a number of rounds which is a multiple of 4, namely AES-192, it means that an additional transformation should be applied at the end of AES-128 and AES-256 to ensure that the internal state is synchronized with the classical representation. Because AES-128 and AES-256 are composed of 10 and 14 rounds respectively, \(F = SR^2(F)\) should be computed to ensure the correctness of the result. This is can be achieved by means of 1 AND and 3 OR instructions plus 2 logical shifts per register, as detailed in Listing 3.

\[\text{SWAPMOVE}(r, r, 0x00000000, 4);\]

**Listing 3:** C code to apply \(SR^2\) on a slice \(r\) according to the representation in Figure 4.

One disadvantage of fixslicing compared to the classical representation is to require four different implementations of the linear layer. While this is not an issue when considering an unrolled implementation, it will increase the code size in a loop-based setting. To mitigate this concern, an interesting tradeoff is to compute \(SR^2\) every two rounds so that
only two different MixColumns implementations are required. We refer to this version as *semi-fixsliced* whereas *fully-fixsliced* refers to a total omission of the ShiftRows. A visual representation is provided in Figure 8.

The Table 1 summarizes the number of operations required for the AES linear layer over 4 rounds, for the fully/semi-fixsliced representations. When considering the overall AES-128 algorithm, the linear layer (by processing 2 blocks at a time) requires 1907 and 915 operations for the classical bitsliced and fixsliced representations, respectively. While this corresponds to an improvement of 52%, the gain might even be more important on some platforms since both representations respectively include 1283 and 563 logical operations, which means an improvement of 56% for this kind of instructions (which are the ones that really matter on ARM). Practical implementations results on ARM Cortex-M and E31 RISC-V processors are reported in the next section.

**Table 1:** Number of operations required to compute the AES linear layer over 4 rounds when processing 2 blocks in parallel, for different representations. LOP, LSH and ROT refer to logical operations, logical shifts and rotations, respectively.

<table>
<thead>
<tr>
<th>Representation</th>
<th>Ref</th>
<th>Round 0</th>
<th>Round 1</th>
<th>Round 2</th>
<th>Round 3</th>
<th>Total over 4 rounds</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classical bitsliced</td>
<td>[SS16]</td>
<td>131</td>
<td>48</td>
<td>16</td>
<td>131</td>
<td>48</td>
</tr>
<tr>
<td>Fully-fixsliced</td>
<td>Ours</td>
<td>75</td>
<td>32</td>
<td>16</td>
<td>51</td>
<td>16</td>
</tr>
<tr>
<td>Semi-fixsliced</td>
<td>Ours</td>
<td>75</td>
<td>32</td>
<td>16</td>
<td>59</td>
<td>16</td>
</tr>
</tbody>
</table>

### 4.2 Application to the key expansion

As previously mentioned, another requirement of the fixslicing technique is to adapt the round keys so that the bits are properly aligned to ensure the correctness of the AddRoundKey operation. Therefore, the key expansion of the fixsliced AES will inevitably bear the cost of some additional computations. To the best of our knowledge, there is no result reported in the literature for a 32-bit implementation of the AES key schedule in a truly bitsliced manner. Actually, the results reported in [SS16, Sto19] are obtained by computing the AES key schedule using a lookup table (LUT) for the SubBytes before packing the round keys to match the bitsliced representation, resulting in key-dependent memory accesses. However, mounting a cache-timing attack against the key schedule seems unpractical since it is often computed only once per key and such attacks require the key-related index to interact with known variable data over multiple samples. On the other hand, when considering power side-channel attacks, countermeasures should not be only integrated to the round function but also to the key schedule as it constitutes another attack vector. This was actually highlighted by the CHES 2018 side-channel contest, where a masked AES implementation was defeated due to a lack of masking in the key schedule [GJS19]. As a result, we consider two variants: (1) LUT-based to provide a fast key schedule implementation when power side-channel attacks are not a concern and to compare with previous works, (2) truly bitsliced implementation that packs the master key at the beginning before operating on the bitsliced representation through the entire
key expansion. The main advantage of the second variant will be to make the integration of Boolean masking easier.

For the LUT-based key schedule, the overhead introduced by fixslicing will be low since it allows to compute $\text{SR}^{-1}$ for $i \in \{1, 2, 3\}$ on the round keys in a non-bitsliced fashion. It is indeed way more efficient as highlighted by Listing 4. Overall, fixslicing introduces on overhead of 8 logical operations per $\text{SR}^{-2}$ computation and 28 logical operations per $\text{SR}^{-i}$ computations for $i \in \{1, 3\}$, which corresponds to $28 \times 2 + 8 = 64$ and $28 \times 2 = 56$ additional operations per quadruple round for the fully-fixsliced and semi-fixsliced representations, respectively. On the other hand, for a truly bitsliced key expansion, one has to pay an extra cost of 64 logical operations on top of 32 logical shifts per $\text{SR}^{-i}$ computations for $i \in \{1, 3\}$ and 32 logical operations plus 16 logical shifts per $\text{SR}^{-2}$ computations, as previously discussed.

```
/* rk[i] refers to the i-th column of the internal state */
1  t = (rk[0] ^ rk[2]) & 0xff00ff00;
2  rk[0] = rk[0] ^ t;
4  t = (rk[1] ^ rk[3]) & 0xff00ff00;
```

Listing 4: C code to apply $\text{SR}^2$ on a round key $\text{rk}$ in a non-bitsliced fashion.

5 Implementation results

While the previous section has shown that fixsliced AES should outperform the current best results on 32-bit platforms, practical implementations are necessary to support our claim. Although the number of operations in the linear layer are reduced by 52% in
theory, it may not lead to the same result when put into practice. For instance, the number of general-purpose registers on a given platform might be too small to contain all the working variables without paying extra memory accesses, or additional cycles might be required to load the bitmasks used in the byte-wise rotations. This section reports implementation results on ARM Cortex-M and E31 RISC-V processors for all the new representations introduced above, in order to practically assess the relevance of fixslicing the AES. All implementations come in two variants: (1) fully unrolled to achieve the best speed results and to compare with previous works, (2) non-unrolled with limited impact on code size. Note that the second variant does not intend to achieve the smallest possible implementation results, but to provide an efficient tradeoff which is more realistic with practical deployments in mind. For our benchmarks, we simply measure the clock cycles spent by one function call. Note that our AES encryption routines process two blocks in parallel without any mode of operation. This choice was mainly motivated to make our implementations malleable in the sense that they can be easily adapted to match any mode of operation. On the other hand, the results reported in [SS16, Sto19] that we use for comparative purposes were obtained by averaging on the processing of 4096 bytes in CTR mode. While our benchmarks do not measure the small overhead due to the CTR mode (which consists in loading the plaintext, performing an XOR with the keystream and storing the result back), the average over 256 blocks cancels the function call overhead (which includes the cycles required to store/restore the context at the beginning and the end of the function) because their AES implementation is fully inlined in the CTR encryption function. All in all, we believe our comparison is fair and might even be slightly in favor of previous works. Our implementations are publicly available at: https://github.com/aadomn/aes.

5.1 ARM Cortex-M

The ARM Cortex-M family refers to 32-bit ARM processors with different computational capabilities. They are all composed of 16 32-bit registers from which two of them (i.e. the program counter and the stack pointer) cannot be freely used, leaving 14 registers available for general use. Bitwise and arithmetic operations (e.g. XOR, AND, OR) require 1 cycle while memory accesses require \( n + 1 \) cycles, where \( n \) is the number of registers to load/store. A very appreciable feature of ARM processors is the inline barrel shifter, which allows combining a logical or circular shift with an arithmetic or bitwise operation at zero cost. Our AES assembly implementations have been benchmarked on Cortex-M3 and Cortex-M4 processors using the STM32L100C and STM32F407VG development boards. Regarding the non-linear layer, the smallest known circuit of the AES S-box consists of 113 gates [BP10, Cal16]. However, because it uses numerous temporary variables, it is not possible to directly implement it using 113 instructions on ARM. Thanks to an ARM-specific instruction scheduler [Sto16], Schwabe and Stoffelen were able to achieve a bitsliced implementation of the SubBytes using 32 additional memory accesses (16 loads and 16 stores) [SS16]. As we did not manage to improve this result, our ARM implementations use the exact same code for this part of the algorithm. When it comes to fixsliced MixColumns, one has to manipulate bitmasks at some point in order to compute the byte-wise rotations. On ARM, by combining the barrel shifter with the BIC instruction, which corresponds to an AND where a NOT is applied to the second operand, it is possible to implement all four fixsliced MixColumns with a single mask and without any memory access. Therefore, the only overhead is the setting of the appropriate mask value in a register, which can be done in 2 cycles on ARM using the MOVW and MOVT instructions. Results are reported in Table 2, where emboldened and italic fonts refer to unrolled and non-unrolled variants, respectively. Note that for the non-unrolled bitsliced implementations of the key schedule, we do not include the code size of the SubBytes and the packing routine, as it is already included in the AES encryption benchmark.
Table 2: Implementation results on ARM Cortex-M3 and M4 for various bitsliced representations of AES. For encryption routines, speed is expressed in cycles per block and the RAM requirements for the round keys are enclosed in parentheses. Emboldened and italic fonts refer to unrolled and non-unrolled implementations, respectively.

<table>
<thead>
<tr>
<th>Representation</th>
<th>Ref</th>
<th>Parallel</th>
<th>Speed (cycles)</th>
<th>ROM (bytes)</th>
<th>RAM (bytes)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Instances</td>
<td></td>
<td>M3</td>
<td>M4</td>
</tr>
<tr>
<td><strong>AES-128 key expansion (LUT-based)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bitsliced</td>
<td>[SS16]</td>
<td>1</td>
<td>1028</td>
<td>1034</td>
<td>3384</td>
</tr>
<tr>
<td>Semi-fixsliced</td>
<td></td>
<td>1</td>
<td>1158</td>
<td>1235</td>
<td>3768</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 461</td>
<td>1 538</td>
<td>784</td>
</tr>
<tr>
<td>Fully-fixsliced</td>
<td>Ours</td>
<td>1</td>
<td>1 178</td>
<td>1 255</td>
<td>3 848</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 481</td>
<td>1 561</td>
<td>936</td>
</tr>
<tr>
<td>Barrel-shiftrows</td>
<td></td>
<td>1</td>
<td>2 406</td>
<td>2 479</td>
<td>7 476</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2 877</td>
<td>2 956</td>
<td>684</td>
</tr>
<tr>
<td><strong>AES-128 key expansion (fully bitsliced)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Semi-fixsliced</td>
<td></td>
<td>2</td>
<td>3 425</td>
<td>3 430</td>
<td>12 560</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3 714</td>
<td>3 7 45</td>
<td>808</td>
</tr>
<tr>
<td>Fully-fixsliced</td>
<td>Ours</td>
<td>2</td>
<td>3 533</td>
<td>3 538</td>
<td>12 928</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3 896</td>
<td>3 939</td>
<td>962</td>
</tr>
<tr>
<td><strong>AES-128 encryption</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bitsliced</td>
<td>[SS16]</td>
<td>2</td>
<td>1 617</td>
<td>1 618</td>
<td>12 120</td>
</tr>
<tr>
<td>Semi-fixsliced</td>
<td></td>
<td>2</td>
<td>1 312</td>
<td>1 314</td>
<td>9 504</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 394</td>
<td>1 415</td>
<td>1 936</td>
</tr>
<tr>
<td>Fully-fixsliced</td>
<td>Ours</td>
<td>2</td>
<td>1 273</td>
<td>1 275</td>
<td>9 184</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 348</td>
<td>1 369</td>
<td>2 360</td>
</tr>
<tr>
<td>Barrel-shiftrows</td>
<td></td>
<td>8</td>
<td>1 289</td>
<td>1 289</td>
<td>37 064</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 517</td>
<td>1 524</td>
<td>2 532</td>
</tr>
<tr>
<td><strong>AES-256 encryption</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Semi-fixsliced</td>
<td></td>
<td>2</td>
<td>1 804</td>
<td>1 806</td>
<td>13 136</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 918</td>
<td>1 945</td>
<td>1 976</td>
</tr>
<tr>
<td>Fully-fixsliced</td>
<td>Ours</td>
<td>2</td>
<td>1 745</td>
<td>1 747</td>
<td>12 656</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 849</td>
<td>1 874</td>
<td>2 392</td>
</tr>
<tr>
<td>Barrel-shiftrows</td>
<td></td>
<td>8</td>
<td>1 677</td>
<td>1 677</td>
<td>47 960</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2 047</td>
<td>2 055</td>
<td>2 532</td>
</tr>
</tbody>
</table>
5.2 RV32I

RISC-V is an open source standard instruction set architecture (ISA) free to use by anyone for any application. The base ISA refers to the minimal set of capabilities any RISC-V core has to implement. The base ISA for 32-bit and 64-bit architectures, namely RV32I and RV64I, are now finalized while a 128-bit and a smaller 32-bit variants are still under development. Among the 32 32-bit registers in RV32I, up to 31 of them are available for general use. This can be a significant advantage over the ARM architecture for algorithms that require many temporary variables. On the other hand, the base ISA is smaller with 21 arithmetic/logic instructions. Note that while logical shifts are available, there is no rotate instruction. However it will be possible to implement it thanks to the BitManip extension [Wol20], which is still under development at time of writing. Indeed, the base ISA can be extended by means of standard extensions, but it comes at a cost in terms of manufacturing and engineering. Cryptographic instruction set extensions for RISC-V actually constitute an active research topic, especially for the AES block cipher [San20, MNP+20]. Our RISC-V implementations rely on the RV32I base ISA, without the use of any extension. For our benchmark, we used the HiFive1 Rev B development board which includes a 32-bit E31 RISC-V core. Bear in mind that the base ISA does not specify the cycles required for each instruction as it depends on the CPU design, therefore the results may vary across RISC-V boards. Our benchmark results are reported in Table 3. Note that for some fully unrolled implementations, the results are omitted because the code size was too large to fit the 2-way instruction cache of 16KiB, resulting in inconsistent measurements.

5.3 Interpretation and discussions

While the barrel-shiftrows representation is not relevant on ARM due to the limited number of general-purpose registers available, it fits very well the RV32I architecture, improving the previous results reported on this platform by 36% with 79 cpb. Note that among the 79 cpb, about 8 are spent to pack/unpack the data into the bitsliced representation. Indeed, the packing routine introduces a significant overhead since there are \(8 \times 128 = 1024\) bits to rearrange in order to match the barrel-shiftrows representation. Therefore, results can be further improved by considering a version of AES that considers that the input data is directly coming in the appropriate format. Since this is basically a matter of perspective, it does not affect the security and this approach was actually adopted to enhance the software performance of the GIFT-COFB authenticated encryption scheme [BCI+20]. Although the barrel-shiftrows representation has considerable RAM requirements, it might be of interest on RV32I platforms for use-cases that have to deal with a large amount of data (e.g. a firmware update). However, it is not well suited for ARM and fixslicing is more relevant on this architecture with 80 cpb in the unrolled setting, which is 21% faster than the classical bitsliced approach. The results are also convincing on E31 with an improvement of 26%. Still, as already mentionned in Section 2.2, the results previously reported that we are comparing to are not optimal since their ShiftRows implementations can be further optimized. In order to fairly evaluate the advantages of fixslicing over naive bitslicing in the case of AES, we give hereafter estimates on how naive bitsliced AES implementations would benefit from the optimization described in Listing 2. Since the SWAPMOVE technique can be implemented using 1 AND, 3 XOR and 2 shifts instructions, it means that the entire ShiftRows can be computed using 64 1-cycle instructions on ARM on top of 4 cycles to load the two corresponding masks, which is an improvement of \(104 - 68 = 36\) cycles per round. All in all, we expect optimized naive bitsliced AES-128 to run around 92 cpb which means that the fully-fixsliced variant would be still faster by 13% on ARM Cortex-M. Regarding the RV32I architecture, the ShiftRows optimization is more valuable since it allows to save \((19 - 12) \times 8 = 56\) cycles per round. Therefore, we expect optimized naive
Table 3: Implementation results on E31 RISC-V processor for various bitsliced representations of AES. For encryption routines, speed is expressed in cycles per block and the RAM requirements for the round keys are enclosed in parentheses. Emboldened and italic fonts refer to unrolled and non-unrolled implementations, respectively.

<table>
<thead>
<tr>
<th>Representation</th>
<th>Ref</th>
<th>Instances</th>
<th>Speed (cycles)</th>
<th>ROM (bytes)</th>
<th>RAM (bytes)</th>
<th>Code</th>
<th>Data</th>
<th>In/Output</th>
<th>Stack</th>
</tr>
</thead>
<tbody>
<tr>
<td>AES-128 key expansion (LUT-based)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bitsliced</td>
<td>[Sto19]</td>
<td>1</td>
<td>1 239</td>
<td>4 736</td>
<td>1024</td>
<td>368</td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Semi-fixsliced</td>
<td>1</td>
<td>1 435</td>
<td>5 024</td>
<td>1024</td>
<td>368</td>
<td>48</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fully-fixsliced</td>
<td>Ours</td>
<td>1</td>
<td>1 464</td>
<td>5 136</td>
<td>1024</td>
<td>368</td>
<td>48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Barrel-shiftrows</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AES-128 key expansion (fully bitsliced)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Semi-fixsliced</td>
<td>2</td>
<td>3 598</td>
<td>2 416</td>
<td>0</td>
<td>368</td>
<td>64</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fully-fixsliced</td>
<td>Ours</td>
<td>2</td>
<td>3 697</td>
<td>2 932</td>
<td>0</td>
<td>368</td>
<td>64</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AES-128 encryption</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bitsliced</td>
<td>[Sto19]</td>
<td>2</td>
<td>1 990</td>
<td>13 208</td>
<td>0</td>
<td>32 (+352)</td>
<td>64</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Semi-fixsliced</td>
<td>2</td>
<td>1 524</td>
<td>11 564</td>
<td>0</td>
<td>32 (+352)</td>
<td>72</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fully-fixsliced</td>
<td>Ours</td>
<td>2</td>
<td>1 468</td>
<td>11 144</td>
<td>0</td>
<td>32 (+352)</td>
<td>72</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Barrel-shiftrows</td>
<td></td>
<td>8</td>
<td>1 263</td>
<td>2 800</td>
<td>0</td>
<td>128 (+1 408)</td>
<td>192</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AES-256 encryption</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Semi-fixsliced</td>
<td>2</td>
<td>2 172</td>
<td>2 602</td>
<td>0</td>
<td>32 (+480)</td>
<td>72</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fully-fixsliced</td>
<td>Ours</td>
<td>2</td>
<td>2 072</td>
<td>3 016</td>
<td>0</td>
<td>32 (+480)</td>
<td>72</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Barrel-shiftrows</td>
<td></td>
<td>8</td>
<td>1 691</td>
<td>2 800</td>
<td>0</td>
<td>128 (+1 920)</td>
<td>192</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
bitsliced AES-128 to run at 106 cpb on E31 processors, which decreases the gain of the fully-fixsliced variant from 26% to 14% on this platform.

Regarding the key schedule, as expected, our LUT-based implementations are all slower than the one previously reported on both platforms. However, we think that it does not call into question the relevance of our results. First, it may be possible to compute the key schedule only once per key before storing all the round keys in (non-volatile) memory if there is enough space available. Second, for each of our implementations, the encryption efficiency takes precedence over the key expansion overhead even when considering only the minimum number of blocks to process. Although there is no previous work of fully bitsliced key schedules on those platforms, we do not think the classical bitsliced representation would be significantly advantaged since the AES key expansion is intrinsically not well suited for bitslicing. Indeed, as reported in Tables 2 and 3, one can observe an overhead factor of about 3 in terms of performance when compared to the LUT-based implementations. On the other hand, note that it allows us to expand two different keys at the same time which means that the number of cycles is divided by a factor of 2 in this case. The ineffectiveness of the bitsliced key schedule is mainly due to the fact that, as illustrated in Figure 2, the S-box is only applied to a single column which means that in a bitsliced setting, the other three columns are updated for nothing. This is the reason why we do not report results for a fully bitsliced key schedule to match the barrel-shiftrows representation, as the overhead would have been too important. Therefore, it implies that when power side-channel attacks constitute a threat, the barrel-shiftrows representation should not fit the needs since the key schedule will be very costly, without even mentioning the RAM requirements.

5.4 Taking first-order masking into consideration

Since the introduction of Boolean masking as a generic countermeasure against power side-channel attacks [CJRR99], many works have been undertaken to assess its impact when applied to the AES. The basic principle is to split each intermediate variable \( x \) into \( d + 1 \) random shares, where \( d \) is called the masking order, such that their sum equals the protected value (i.e. \( x = x_0 \oplus x_1 \oplus \cdots \oplus x_d \)). The higher the masking order, the more difficult it is to practically defeat a cryptographic implementation. In this section, we only focus on first-order masking schemes.

Regarding software implementations on ARM, the best results reported in the literature shows that one should expect a penalty factor of around 5 in terms of performance [BGRV15, SS16]. Note that this includes the generation of randomness, which is highly platform dependant and can constitute a real burden for the most resource-constrained devices. To tackle this issue, a first-order masking scheme that requires only two random bits per block has recently been published [GSDM+19]. Their masking scheme requires that all bytes within the internal state are masked by the following random value

\[
m_1 \| m_0 \oplus m_1 \| m_0 \oplus m_1 \| m_0 \| m_0 \| m_0 \| m_1 \| m_0 \| m_1
\]

where \( m_0, m_1 \) refer to the two random bits and \( \| \) refers to bit concatenation. On top of reducing the amount of randomness to generate, this scheme allows to achieve very competitive performance. Usually, first-order masked implementations slow down the runtime of the linear layer by a factor 2 since it has to be computed on both shares. In this scheme however, because the mask remains the same through the entire AES encryption, one has just to remask some variables to ensure that no values with the same mask get combined. Moreover the SubBytes can be efficiently implemented using a dedicated AND gate. All in all, their implementation runs at 212 cpb, which is the fastest first-order AES implementation reported on ARM Cortex-M4 at the time of writing. Because this result was achieved using the classical bitsliced representation detailed in Figure 4, we
can easily adapt their implementation to match the fixsliced representation. We run our benchmark on the ARM Cortex-M4 only as this is the only one that embeds a random number generator among our three development boards. Note that this is the same board as the one used in [GSDM+19]. Our benchmark results are reported in Table 4.

Table 4: First-order masked implementation results on ARM-Cortex M4 for various bitsliced representations of AES-128. For encryption routines, speed is expressed in cycles per block and the RAM requirements for the round keys are enclosed in parentheses.

<table>
<thead>
<tr>
<th>Representation</th>
<th>Ref</th>
<th>Parallel Instances</th>
<th>Random (bits)</th>
<th>Speed (cycles)</th>
<th>ROM (bytes)</th>
<th>RAM (bytes)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Code I/O Stack</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>AES-128 key expansion (fully bitsliced)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Semi-fixsliced</td>
<td>2</td>
<td>2</td>
<td>7178</td>
<td>6576</td>
<td>401</td>
<td>200</td>
</tr>
<tr>
<td>Fully-fixsliced</td>
<td>2</td>
<td>2</td>
<td>7317</td>
<td>27448</td>
<td>401</td>
<td>200</td>
</tr>
<tr>
<td><strong>AES-128 encryption</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bitsliced</td>
<td>[GSDM+19]</td>
<td>2</td>
<td>3388</td>
<td>25200</td>
<td>32 (+352)</td>
<td>188</td>
</tr>
<tr>
<td>Semi-fixsliced</td>
<td>Ours</td>
<td>2</td>
<td>3055</td>
<td>23754</td>
<td>32 (+401)</td>
<td>188</td>
</tr>
<tr>
<td>Fully-fixsliced</td>
<td></td>
<td>2</td>
<td>2989</td>
<td>22086</td>
<td>32 (+401)</td>
<td>188</td>
</tr>
</tbody>
</table>

Because 2 blocks are processed in parallel, 4 random bits are generated per encryption routine. More precisely, the 3 32-bit masks $M_0, M_1, M_2$ are defined in the following way:

\[
M_0 = m_0 || m'_0 || \cdots || m_0 || m'_0 \\
M_1 = m_1 || m'_1 || \cdots || m_1 || m'_1 \\
M_2 = m_0 \oplus m_1 || m'_0 \oplus m'_1 || \cdots || m_0 \oplus m_1 || m'_0 \oplus m'_1
\]  

such that 2 different random bits are used for every block. For our masked key expansion, because our implementations allow to pass two different keys as parameters, 4 random bits would be sufficient as well. However, because our benchmarking platform generates 32-bit random words, we decided to mask each round key with a different key since it only requires to generate an additional 32-bit random word. Therefore, our masked AES-128 key schedule requires 44 random bits in total. Once again, the performance results for the key expansion are given by considering that the same key is used to encrypt both blocks, and the results can be halved if two different keys are used. Regarding encryption routines, we observe a performance gain of up to 12% thanks to the fixslicing technique. Note however that since the round keys use different masks, we are able to save some XOR instructions to do some remasking in the AddRoundKey. The fact that the improvement is less significant for the first-order masked implementations is mainly due to the masking scheme. Indeed, since each byte is masked using the same bits, the ShiftRows is only computed once since there is no need to adjust the masks accordingly. Moreover the MixColumns only bears the cost of some additional XOR instructions for remasking purposes. Therefore, we expect our fixsliced representations to be even more of interest for other masking schemes that do not rely on the same masks for all bytes and requires to compute the linear layer on both shares.

However, because the practical security of an implementation depends on numerous factors, other first-order masked AES-128 implementation results reported in the literature
may offer a better security guarantee at the cost of a lower throughput. Therefore, benchmarks of masked implementations should be considered with caution since security parameters have to be taken into account. In the case of [GSDM+19], as pointed out by the authors, it is very likely that the reuse of randomness in their masking scheme may introduce some weaknesses (e.g. an increase of the signal-to-noise ratio) that could facilitate an attack in practice. We emphasize that our goal was mainly to highlight that fixslicing allows us to improve the fastest masked AES-128 implementation reported at the time writing, even though the corresponding masking scheme has a low impact on the linear layer.

6 Application to another AES-like design: Skinny

The lightweight family of tweakable block ciphers Skinny [BJK+16] has two block versions: 64-bit and 128-bit. Hereafter, we only consider the case of Skinny-128 for consistency with our work on AES described above. Like AES, the internal state of Skinny-128 consists of a $4 \times 4$ square array of bytes. One encryption round is composed of five operations in the following order: SubBytes, AddConstants, AddRoundTweakey, ShiftRows and MixColumns as illustrated in Figure 9. While Skinny shows outstanding results when implemented in hardware, the picture is more mixed when it comes to software. Although its original publication reports bitsliced implementations of Skinny-128-128 that reach 3.78 and 3.43 cpb on Haswell and Skylake architectures respectively, they rely on the Intel AVX2 instruction set and require to process 64 blocks in parallel. To date, it is not very clear how Skinny performs on 32-bit microcontrollers since the only dedicated implementations publicly available are the ones from Weatherley [Wea17]. His implementations are byte-sliced in the sense that each row of the internal state is represented by a 32-bit word. Therefore, the ShiftRows and the MixColumns simply consist of 3 32-bit rotations and 3 exclusive-ORs, respectively. On the downside, this representation requires to apply many masks and shifts to compute the SubBytes in a constant-time manner. More precisely, it requires 28 logical operations and 20 logical shifts per word. In the following, we consider a bitsliced approach and detail the benefits of fixslicing in the case of Skinny-128.

![Figure 9: The Skinny round function (from [Jea16])]
the column-wise representation detailed in Figure 10, we are able to adjust the different MixColumns implementations by simply adding some rotations and adjusting the masks. The Listing 6 shows how to compute the MixColumns when the state is synchronized with the classical representation (i.e. \( F = S \)), whereas the Listing 5 considers that \( F = SR^{-1}(S) \). For the two other functions, the same principle applies and the only differences lie in the masks and the rotation values.

\[
\begin{array}{cccccccc}
\text{column 3} & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\text{row 0} & \text{row 1} & \text{row 2} & \text{row 3} & \text{row 0} & \text{row 1} & \text{row 2} & \text{row 3} \\
\hline
R_0 & \text{block} & \text{block} & \text{block} & \text{block} & \text{block} & \text{block} & \text{block} \\
& b_{23} & b_{24} & b_{25} & b_{26} & b_{27} & b_{28} & b_{29} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
R_7 & b_{31} & b_{32} & b_{33} & b_{34} & b_{35} & b_{36} & b_{37} \\
& b_{56} & b_{57} & b_{58} & b_{59} & b_{60} & b_{61} & b_{62} \\
\end{array}
\]

Figure 10: Bitsliced representation for Skinny-128 using 8 32-bit registers \( R_0, \ldots, R_7 \) to process 2 blocks \( b^0, b^1 \) in parallel where \( b_{ij} \) refers to the \( i \)-th bit of the \( j \)-th block.

\[
\begin{align*}
1 & \quad t = r \oplus 0x03030303; \\
2 & \quad r = r \ominus \text{ROR}(t, 30); \\
3 & \quad t = r \oplus 0x03030303; \\
4 & \quad r = r \ominus \text{ROR}(t, 4); \\
5 & \quad t = r \oplus 0x03030303; \\
6 & \quad r = r \ominus \text{ROR}(t, 26); \\
\end{align*}
\]

Listing 5: C code to compute the Skinny-128 MixColumns on a slice \( r \) according to the representation in Figure 10 when \( F = S \).

\[
\begin{align*}
1 & \quad t = \text{ROR}(r, 24) \oplus 0x0c0c0c0c0c; \\
2 & \quad r = r \ominus \text{ROR}(t, 30); \\
3 & \quad t = \text{ROR}(r, 16) \oplus 0x0c0c0c0c0c; \\
4 & \quad r = r \ominus \text{ROR}(t, 4); \\
5 & \quad t = \text{ROR}(r, 8) \oplus 0x0c0c0c0c0c; \\
6 & \quad r = r \ominus \text{ROR}(t, 2); \\
\end{align*}
\]

Listing 6: C code to compute the Skinny-128 MixColumns on a slice \( r \) according to the representation in Figure 10 when \( F = SR^{-1}(S) \).

Therefore, the overhead on the MixColumns introduced by fixslicing is less important in the case of Skinny with only 17 circular shifts over 4 rounds. On ARM, no extra cycles are spent for the rotations and therefore the gain directly corresponds to the cost of the ShiftRows, namely 104 cycles per round. Note that unlike for the AES, a full resynchronization of state occurs every 8 rounds instead of 4, since we also omit the row permutation in the MixColumns. While this is not an issue for the linear layer, it requires 8 different SubBytes implementations to avoid slice renaming. Instead of relying on octuple rounds which would consume a considerable amount of code size, we suggest to rename the slices every four rounds. After 4 rounds, one has simply to swap slices 0 with 1, 2 with 3, 4 with 7, and finally 5 with 6. This can be done using 12 cycles, resulting in an overhead of 3 cycles per round. Our implementations are based on quadruple rounds thanks to this tradeoff. Results for fully-fixsliced implementations of the Skinny-128 family of tweakable block ciphers on ARM-Cortex-M3/4 are reported in Table 5.

Note that we report two implementation versions for each algorithm: one that operates on a single block at a time and another one that processes 2 blocks in parallel. The first variant is possible thanks to some symmetry in the Skinny-128 S-box which allows an efficient computation in a bitsliced manner using only 4 slices instead of 8. More details are given in Appendix A. For comparison purposes, we benchmark the byte-sliced implementations of Skinny-128 that are publicly available. Note however that they are written in C while ours are written in assembly. One can see that our fully-bitsliced implementations are up to 4 and 2.5 times faster when processing 2 and 1 block at a time, respectively. Still, the bitsliced approach increases considerably the amount of RAM to store all the round tweak keys. This could be addressed by computing the tweakkey schedule on the fly, at the expense of performance degradation.
Table 5: Implementation results on ARM Cortex-M3 and M4 for fully-fixsliced implementations of the Skinny-128 family of tweakable block ciphers. Speed is expressed in cycles per block and the RAM requirements for the round keys are enclosed in parentheses.

<table>
<thead>
<tr>
<th>Representation Ref</th>
<th>Parallel Blocks</th>
<th>Speed (cycles)</th>
<th>Code size (bytes)</th>
<th>RAM (bytes)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>M3</td>
<td>M4</td>
<td>I/O</td>
</tr>
<tr>
<td>Skinny-128-128 encryption</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Byte-sliced [Wea17]</td>
<td>1</td>
<td>8 187</td>
<td>8 237</td>
<td>820</td>
</tr>
<tr>
<td>Fully-fixsliced Ours</td>
<td>1</td>
<td>3 055</td>
<td>3 066</td>
<td>1 504</td>
</tr>
<tr>
<td>Fully-fixsliced Ours</td>
<td>2</td>
<td>1 862</td>
<td>1 872</td>
<td>1 620</td>
</tr>
<tr>
<td>Skinny-128-256 encryption</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Byte-sliced [Wea17]</td>
<td>1</td>
<td>9 809</td>
<td>9 867</td>
<td>820</td>
</tr>
<tr>
<td>Fully-fixsliced Ours</td>
<td>1</td>
<td>3 639</td>
<td>3 654</td>
<td>1 520</td>
</tr>
<tr>
<td>Fully-fixsliced Ours</td>
<td>2</td>
<td>2 214</td>
<td>2 224</td>
<td>1 628</td>
</tr>
<tr>
<td>Skinny-128-384 encryption</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Byte-sliced [Wea17]</td>
<td>1</td>
<td>11 433</td>
<td>11 499</td>
<td>820</td>
</tr>
<tr>
<td>Fully-fixsliced Ours</td>
<td>1</td>
<td>4 223</td>
<td>4 238</td>
<td>1 536</td>
</tr>
<tr>
<td>Fully-fixsliced Ours</td>
<td>2</td>
<td>2 566</td>
<td>2 579</td>
<td>1 636</td>
</tr>
</tbody>
</table>

7 Conclusion

In this article, we pushed bitsliced AES to its limits on 32-bit platforms by minimizing the cost of the ShiftRows operation. To do so, we first proposed a new bitsliced representation called barrel-shiftrows that allows to compute the ShiftRows using only 32-bit rotations without impacting the efficiency of the MixColumns. Thanks to this representation, we report that it is possible to reach 81 and 79 cpb for AES-128 (by assuming pre-computed round keys) on ARM Cortex-M and E31 RISC-V processors respectively, smashing the previous results on those platforms by 20% and 36%. On the downside, this representation requires to process 8 blocks in parallel and 1408 bytes to store all the AES-128 pre-computed round keys. In order to come up with an implementation that is more appropriate to resource-constrained devices, we applied the concept of fixslicing to the AES and shown that a total omission of the ShiftRows allows to reduce the number of operations spent by the linear layer over 4 rounds by 52%. Because completely omitting the ShiftRows requires 4 different implementations of the MixColumns, we proposed the semi-fixsliced variant that computes the ShiftRows every 2 rounds, allowing many implementation tradeoffs. Our 32-bit fixsliced AES implementations operate on 2 blocks at a time and require 352 bytes to store all the pre-computed round keys. Overall, we reported that fixsliced AES allows to reach 80 and 91 cpb on ARM Cortex-M and E31 respectively, improving the previous results on those platforms by 21% and 26%. Although the naive bitsliced results previously reported are not optimal due to their ShiftRows implementation, we introduced estimates to show that our implementations should remain the fastest on those platforms. We also applied fixslicing to the fastest first-order masked AES implementation reported in the literature on ARM Cortex-M4 and improved its performance by 12%. As future work, it would be interesting to investigate the benefits of fixsliced AES for other masking schemes, especially at higher orders.
Finally, we demonstrated the genericity of fixslicing in the case of AES-like ciphers by illustrating its use on the Skinny-128 family of tweakable block ciphers, enhancing the performance up to a factor of 4 when compared to the previous implementations reported on 32-bit microcontrollers. More generally, it is very likely that the fixslicing technique might be of interest for other constructions. While this work only focused on 32-bit platforms, fixslicing might lead to improvements on other architectures as well. For instance, even for CPUs featuring particular vector shuffle instructions (e.g. Intel’s SSSE3 pshufb or ARM’s NEON vtbl), adopting a fixsliced approach by using those instructions on temporary variables only (so that the slices remain fixed) could allow saving some instructions as soon as a resynchronization occurs.

Acknowledgements

The authors would like to thank the anonymous reviewers for their helpful comments. The authors are supported by a Temasek Labs grant (DSOCL16194) and a joint WASP/NTU grant.

References


A Fixslicing a single block for Skinny-128

Since the 8-bit S-box used in Skinny-128 is built from the composition of two 4-bit Sboxes, one can remark some symmetry in its definition. For instance, as illustrated in Figure 11, the first four gates can be split in two groups that can be computed in parallel (i.e. \((b_0 \land b_1)\) and \(b_2 \oplus b_3\) on one hand, and \((b_4 \land b_5)\) and \(b_6 \oplus b_7\) on the other hand). From an implementation point of view, it means that we can compute two gates instead of four by dividing each byte in two nibbles and to rearrange them in a column-wise manner within the slices. However, some bit permutations are required between the different layers. The Figure 12 illustrates a way to compute the permutations according to the new representation, so that the bits are always correctly arranged within the nibbles to compute the gates in parallel. Note that one does not need to implement the last permutation as it simply consists in swapping the slices, which is free from a bitslicing point of view.
According to our representation, the bit permutations can be simply implemented by means of bit swaps between the nibbles, which can be efficiently computed thanks to the SWAPMOVE routine defined in Listing 7. As a result, implementing the Sbox in a nibble-wise bitsliced manner allow to divide the number of gates by a factor 2, excluding the cost of 6 calls to the SWAPMOVE routine, which is about $6 \times 4 = 24$ cycles on ARM. While it does not seem really worth it at a first glance, when considering an entire 128-bit input block, our new representation now fits into 4 32-bit words instead of 8 compared to the above mentioned bitsliced implementation. This allows to compute twice faster the addition of the round tweakeys and the MixColumns, saving 4 exclusive-OR (plus 4 load from memory) and $6 \times 4 = 24$ operations per round, respectively. All in all, the extra cost from the SWAPMOVE is cancelled by the save on the MixColumns, resulting in a twice less expensive S-box and AddRoundTweakey operations. Moreover, note that our trick is of great interest when considering countermeasures against power side-channel attacks as it reduces by half the number of non-linear gates, which are costly to secure.

Listing 7: C code for the SWAPMOVE routine.