Efficient Hashing using the AES Instruction Set

Joppe W. Bos[1] and Onur Özen[1] and Martijn Stam[2]

[1] Laboratory for Cryptologic Algorithms, EPFL,
Station 14, CH-1015 Lausanne, Switzerland
{joppe.bos, onur.ozen}@epfl.ch
[2] Department of Computer Science, University of Bristol,
Merchant Venturers Building, Woodland Road, Bristol, BS8 1UB, United Kingdom
stam@cs.bris.ac.uk

Abstract. In this work, we provide a software benchmark for a large range of 256-bit blockcipher-based hash functions. We instantiate the underlying blockcipher with AES, which allows us to exploit the recent AES instruction set (AES-NI). Since AES itself only outputs 128 bits, we consider double-block-length constructions, as well as (single-block-length) constructions based on RIJNDAEL-256. Although we primarily target architectures supporting AES-NI, our framework has much broader applications by estimating the performance of these hash functions on any (micro-)architecture given AES-benchmark results. As far as we are aware, this is the first comprehensive performance comparison of multi-block-length hash functions in software.

1 Introduction

Historically, the most popular way of constructing a hash function is to iterate a compression function that itself is based on a blockcipher (this idea dates back to Rabin [63]). This approach has the practical advantage—especially on resource-constrained devices—that only a single primitive is needed to implement two functionalities (namely encrypting and hashing). Moreover, trust in the blockcipher can be conferred to the corresponding hash function. The wisdom of blockcipher-based hashing is still valid today. Indeed, the current cryptographic hash function standard SHA-2 and some of the SHA-3 candidates are, or can be regarded as, blockcipher-based designs. In the 80s, several methods were proposed with an eye towards using the then-standard Data Encryption Standard (DES) as the underlying primitive [52,39,21]. At present, the contemporary Advanced Encryption Standard (AES [53]) is a more obvious choice instead.

A well-studied class of blockcipher-based hash functions are the PGV hash functions (after Preneel, Govaerts and Vandewalle [62]), encompassing Davies–Meyer (DM) and Matyas–Meyer–Oseas (MMO) as special cases. When based on a blockcipher operating on n-bit blocks with k-bit keys, these functions compress k bits per blockcipher call and they output an n-bit digest. The PGV hash functions are simple (low overhead) and are provably secure in the ideal-cipher model [17]. Yet they suffer from one major drawback: in order to achieve an acceptable level of collision resistance, one needs a primitive operating on more than 160 bits [23]. This rules out most existing blockciphers, *including* AES (which operates on 128-bit blocks only).

As a remedy, *double-block-length* and more generally *multi-block-length* compression and hash functions were introduced. These are compression functions outputting an rn-bit digest (for an integer $r \geq 2$, $r = 2$ for the double-block-length case), even though they are based on a primitive operating only on n-bit blocks. The longer digest size opens up the possibility of collision resistance of 2^n time (primitive evaluations) even when using a relatively small primitive. Today, there is truly a wealth of suitable blockcipher-based constructions to choose from and Table 1 gives an overview of the constructions we consider. We do not consider *all* possibilities, for instance we omit versions of GRØSTL, JH or SPONGE based on RIJNDAEL-256. As can be seen, we instantiate the underlying blockcipher with either AES-128, AES-256 or RIJNDAEL-256. The latter option allows us to consider single-block-length constructions achieving a 256-bit digest (using an AES-related primitive).

Our choice of constructions includes several different design ideas and paradigms. For years, most cryptographic hash function designs revolved around the same principle [63,52,21]: the Merkle-Damgård

Table 1. A brief taxonomy of the schemes considered. N_r stands for the number of rounds.

Blockcipher (dimensions)	**Variable-key Constructions**	**Fixed-key Constructions**
AES-128 ($k = 128, n = 128, N_r = 10$)	MDC-2, MJH, PEYRIN ET AL.(I)	LP362
AES-256 ($k = 256, n = 128, N_r = 12$)	ABREAST-DM, HIROSE-DBL, KNUDSEN–PRENEEL, MJH-DOUBLE, QPB-DBL, PEYRIN ET AL.(II)	n.a.
RIJNDAEL-256 ($k = 256, n = 256, N_r = 14$)	DM	LANE*, LUFFA*, LP231, SS

paradigm. In this cascaded mode of operation, the main focus is to construct a secure and efficient compression function; these properties are then inherited by the overall hash function. Later constructions started to deviate from this paradigm, for instance by some form of strengthening [47,12] or by only targeting security in the iteration [17,9].

A more fundamental design shift occurred in the way the blockcipher itself is used. A blockcipher, operating on n bits with a k-bit key, can already be regarded as a compressing primitive itself. This facilitates the transformation into a proper compression function, but a disadvantage of using a blockcipher this way is that it requires frequent re-keying, which tends to be expensive (see Section 2.1 for details). For this reason, there have been substantial efforts in recent years to design permutation-based compression functions. Obviously, given a blockcipher one can construct a permutation by simply fixing the key (we focus on permutations with either $n = 128$ or 256 bits).

While the design and analysis of multi-block-length compression functions have garnered significant attention, the focus in the literature seems squarely at security evaluation and theoretical notions of efficiency (expressed as the ratio of message blocks compressed per blockcipher call). Although the latter is known to give only a coarse indication of real-life efficiency, actual performance benchmarks, in hard- or software, are normally left as future work. (A notable exception is the work by Bogdanov et al. [18], who provide hardware benchmarks for some multi-block-length compression functions in hardware using the lightweight blockcipher PRESENT as the underlying building block.)

Our Contribution. In this work we bring together the mainly theoretical world of compression function designs with the practical demand of fast implementations. Instantiating the blockcipher-based primitives with AES-128, AES-256, respectively RIJNDAEL-256 (and their fixed-key versions to build permutations), we obtain hash functions with a fixed 256-bit digest size. Apart from three constructions (LANE*, LUFFA* and KNUDSEN–PRENEEL) all constructions have known proofs of security in the ideal-cipher model (see Appendix A for a more detailed discussion on the security of our target constructions). The former SHA-3 candidates LANE* and LUFFA* do not have security proofs, neither for collision resistance nor for preimage resistance. We include them in our benchmark (with different building blocks) to illustrate their performance capabilities; KNUDSEN–PRENEEL is another exception where a good collision resistance lower bound is still an open problem.

To the best of our knowledge, this is the first overview of software implementations of the most studied and influential blockcipher- and permutation-based compression and hash functions. The target designs (see also Table 1) have been implemented and measured on an Intel Core i5 650 (3.20GHz) using C intrinsics to implement the various SS(S)E{2,3,4} and the recent AES instruction set (AES-NI) extensions [26,27]. Although measured on a single Intel architecture with AES-NI we expect the relative performance obtained to be representative for other Intel architecture families with AES-NI support as well. The Intel compiler version 12.0.0 and GNU Compiler Collection (gcc) version 4.4.3 were used for

code compilation. For each design we performed specific optimizations to fully exploit AES-NI. The details are discussed in Section 3, with Table 3 providing a summary of our findings.

Our major conclusion (see also Section 4) is that, when assuming that the underlying primitives behave ideally, one can obtain fast and provably secure blockcipher based hash functions on soon to be mainstream architectures supporting AES-NI.

The Choice for AES. Our choice for AES (and RIJNDAEL-256) is a natural one: it is the official US and de facto world standard blockcipher. AES' prime position has led to a large body of research on AES, both on its security and implementation. Consequently, AES runs very fast in hard- and software, making AES an obvious choice from a performance perspective. The deal is sweetened further by the recent introduction of AES-NI. Indeed, as reported in [27], one can achieve significant speed using the new instruction set (e.g. up to 1.3 cycles/byte on a single core Intel Core i7-980X for AES-128 in parallel modes). To benefit from synergy with AES and AES-NI in particular, several SHA-3 candidates were instantiated by using some of AES components as well (e.g. the AES round function), which was later demonstrated to indeed lead to fast hashing [4]. Our goal here is to investigate the potential of AES-NI for fast hashing even further by focusing on well-known blockcipher- and permutation-based (compression function) designs that can be instantiated with AES (or more generally RIJNDAEL).

From a security perspective, AES remains unbroken as a blockcipher in the standard setting. It has survived many years of cryptanalysis and a practical break of this cipher would have a significant impact on the cryptographic landscape. Nonetheless, our choice for AES will not be without detractors as a consequence of recent related-key attacks on AES [14,15]. The theoretical ramifications of a related-key attack to hash-function security are still unclear. Any serious related-key attack undermines the assumption that the blockcipher behaves 'ideally', but this need not lead to any deviant behaviour of the hash function itself (especially if its proof uses the weaker unforgeable-cipher model). Of course, in practice a related-key attack is often underpinned by some other (well-defined) weakness and exploiting this weakness directly (ignoring the derived related-key attack) might be more fruitful when attacking the hash function. For instance, Khovratovich [34, Corollary 2] states unambiguously that "AES-256 in the Davies–Meyer hashing mode leads to an insecure hash function" but later provides solace by remarking that it is not known how the techniques used against AES-256 Davies–Meyer can be modified to attack double-block-length constructions (the focus of this paper).

As a final remark, the timings we obtain evidently depend strongly on the number of rounds used by AES. While one can argue that the number of rounds used should be fine-tuned for each of the hash functions (increasing or decreasing, depending on the perceived security margin), we believe that using AES as is will give the cleanest comparison (and any changes might be considered contentious).

2 Building Blocks

2.1 The Advanced Encryption Standard (AES)

The Blockciphers AES-128, AES-256 and RIJNDAEL-256. AES is a member of the RIJNDAEL blockcipher suite [20,53]. It was standardized by the US National Institute of Standards and Technology (NIST) after a public competition similar to the one currently ongoing for SHA-3 [55]. AES operates on an internal state of 128 bits while supporting 128-, 192-, and 256-bit keys. The internal state is organized in a 4×4 array of 16 bytes, which is transformed by a round function N_r times. The number of rounds is $N_r = 10$ for the 128-bit key, $N_r = 12$ for the 192-bit key, and $N_r = 14$ for the 256-bit key variants. In order to encrypt, the internal state is initialized, then the first 128-bits of the key are `xor`ed into the state, after which the state is modified $N_r - 1$ times according to the round function, followed by a slightly different final round. The round function consists of four steps: `SubBytes`, `ShiftRows`, `MixColumns` and `AddRoundKey` (the final round omits the `MixColumns` step), each of which operates on the 128-bit state (for the exact details see the AES specification [20,53]). The larger state variant

of AES, RIJNDAEL-256, operates almost in the same way except for the ShiftRows operation, a state size of 256 bits, a 256-bit key and $N_r = 14$ rounds.

Nine years after becoming the symmetric encryption standard, the only theoretical attack on the full AES is restricted to the related key scenario and even then applies only to the 192-bit [14] and 256-bit key versions [14,15]. So far no theoretical attacks on all rounds of AES-128 are known. More cryptanalytic success has been achieved by using the characteristics from the actual implementation of AES, for instance cache attacks [57,5] can recover an AES key in only 65 milliseconds (Tromer et al. [76] give a more detailed survey of side-channel attacks against AES). However, side-channel analysis is far less of a concern for hash functions (except for MACs based on hash functions, such as HMAC) and we will blithely ignore the issue in this paper.

The AES Instruction Set (AES-NI). Designing fast implementations of AES, which overcome the various software side channel attacks, has been an active research area in recent years. The latest examples are bitsliced implementations for Intel core i7 architectures [10,33] and implementations which target a variety of common CPU architectures [29,7]. In the latter an overview of the state-of-the-art fast AES implementations is given. The fastest AES implementations targeting microcontrollers are described in [56].

In the last decade, use of the *single instruction, multiple data* (SIMD) paradigm has become a general trend in computer architecture design. It enhances the speed of software implementations by offloading the computational work to special units which operate on larger data types, improving overall throughput. In 1999, Intel introduced the streaming SIMD extensions (SSE), a SIMD instruction set extension to the x86 architecture. One of the latest additions to these extensions is the AES instruction set [26,27] available in the 2010 Intel Core processor family based on the 32nm Intel micro-architecture named Westmere. This instruction set will also be supported by AMD in their next-generation CPU "Bulldozer". (Note that previously several instruction set extensions have been suggested towards improving the performance of AES [74,73,8,75].) AES-NI does not only increase the performance of AES (as well as any version of RIJNDAEL) but also runs in data-independent time and by avoiding the use of any table lookups the aforementioned cache attacks are avoided. This instruction set consists of six new instructions. At the same time, a new instruction for performing carry-less multiplication is released in the CLMUL instruction set extension. We can summarize the new instructions as follows [26,27,28]:

- AESENC and AESDEC perform a single round of encryption, resp. decryption.
- AESENCLAST and AESDECLAST perform the last round of encryption resp. decryption.
- AESKEYGENASSIST is used for generating the round keys used for encryption.
- AESIMC is used for converting the encryption round keys to a form usable for decryption using the Equivalent Inverse Cipher.
- PCLMULQDQ performs carry-less multiplication of two 64-bit operands to an 128-bit output.

Throughput Considerations. Many of the constructions targeted in this paper require the computation of more than one call to a blockcipher (with or without fixed-key). If these two or more calls can be run concurrently (while possibly sharing the key expansion), a performance gain can be expected as AES round instructions are pipelined and can be dispatched theoretically every 1-2 CPU clock cycles, provided that all data is available on time and there is no dependency between such subsequent calls [26]. Since the latency of a single round instruction is 5 cycles [25], running multiple independent blockciphers increases the overall throughput. The same reasoning holds when implementing a single RIJNDAEL-256 component. This sibling of AES works on an internal state of 256 bits and it is implemented using two data-independent calls to AESENC.

In the context of encryption, several performance results of AES exploiting AES-NI have been presented [27,28,49]. These works show that using AES-NI tends to give very fast implementations when

Table 2. Our experimental results on the encryption and key expansion routines for AES-128, AES-256 and RIJNDAEL-256. The entries show the results in cycles per operation and cycles per byte together with the compiler, icc (i) or gcc (g), resulting in the fastest code. In the table **K** and **E** denote the key expansion and the encryption respectively. The upper part of the table shows the results of several independent key expansions and encryption operations that are called in parallel. In the lower part, **xKyE** denotes x independent key schedules followed by y independent encryptions. If $x = 1$ all encryptions use the same expanded key, if $x = y$ all encryptions use a different expanded key. For comparison, the performance details of the Intel AES-NI sample library on our platform are stated as well.

Algorithm	**Operation**															
	1K		**2K**		**3K**		**4K**		**1E**		**2E**		**3E**		**4E**	
AES-128	97.7	6.1 (g)	126.1	3.9 (g)	163.4	3.4 (g)	226.7	3.5 (i)	60.2	3.8 (i)	60.6	1.9 (i)	67.7	1.4 (i)	84.7	1.3 (i)
AES-256	125.5	3.9 (g)	147.2	2.3 (g)	202.6	2.1 (i)	287.2	2.2 (i)	82.0	5.1 (i)	83.0	2.6 (i)	93.6	1.9 (i)	113.9	1.8 (i)
RIJNDAEL-256	291.6	9.1 (g)	316.6	5.0 (g)	412.6	4.3 (g)	570.3	4.5 (i)	182.9	5.7 (i)	219.2	3.4 (g)	281.4	2.9 (i)	352.6	2.8 (g)
	1K1E		**2K2E**		**3K3E**		**4K4E**				**1K2E**		**1K3E**		**1K4E**	
AES-128	107.4	3.4 (g)	149.2	2.3 (g)	200.0	2.1 (g)	269.9	2.1 (g)			120.1	2.5 (g)	135.3	2.1 (g)	137.8	1.7 (g)
AES-256	152.8	3.2 (g)	178.1	1.9 (g)	249.7	1.7 (g)	337.9	1.8 (g)			154.0	2.4 (g)	158.4	2.0 (g)	164.9	1.7 (g)
RIJNDAEL-256	285.3	4.5 (i)	407.5	3.2 (i)	620.5	3.2 (i)	867.3	3.4 (i)			312.0	3.6 (g)	373.3	2.9 (i)	463.7	2.9 (g)

Intel AES-NI Sample Library

Algorithm	**1K**		**1E**	**4E**	**Algorithm**	**1K**		**1E**	**4E**
AES-128	99.0	6.2	4.0	1.3	AES-256	124.5	7.8	5.4	1.7

multiple blockcipher calls can be made in parallel (incidentally, they also show that the optimal way to interleave the instructions is hard to pin down). However, they are of limited use to predict the runtimes of AES-based hash functions as rekeying tends to be far more frequent in the hashing scenario than in the encryption one. Indeed, for blockcipher-based compression functions considered in this paper, the key-scheduling needs to be performed for every compression function evaluation and that results in a significant overhead. For this reason, we start with a detailed performance overview of AES and RIJNDAEL-256 that takes rekeying into account. Table 2 contains performance details when running multiple key expansions, encryptions or a combination of the two. In order to conduct these experiments we created a code generator which, when given a number of x key expansions and y encryptions, tries different strategies to implement these functionalities. The performance numbers presented in Table 2 are an average over millions of data-dependent runs. For comparison, we also included timings from Gueron's hand-crafted assembly code [27,28] as used in the Intel AES-NI sample library. (Note that, roughly speaking, our measure **1E** coincides with AES run in a chaining mode such as CBC or CFB, whereas AES run in a parallel mode such as CTR or ECB is closer to the best time we get for **xE**, see Table 2 for the performance details).

2.2 Finite Field Arithmetic ($\mathbb{F}_{2^m}$ Full/Scalar Multiplication)

Some of the compression function designs we consider require finite field multiplication, in particular in $\mathbb{F}_{2^{128}}$ and $\mathbb{F}_{2^{256}}$. There is some freedom in how to represent the fields—the security proofs for the hash functions are independent of this choice—so we opt for the usual representation of elements in $\mathbb{F}_{2^m}$ as polynomials over $\mathbb{F}_2$ reduced modulo an irreducible polynomial of degree m. We use $x^{128} + x^7 + x^2 + x + 1$ as irreducible polynomial for $m = 128$ and $x^{256} + x^{10} + x^5 + x^2 + 1$ for $m = 256$.

Multiplication in $\mathbb{F}_{2^{128}}$ is implemented using the code examples as described in [28] in the setting of implementing the Galois counter mode. This is realized by using the new instruction `PCLMULQDQ` to implement the multiplication; this instruction calculates the carry-less product of the two 64-bit input to an 128-bit output. Note that this instruction has a latency of 12 cycles and can be dispatched every 8 cycles [25]. Hence, compared to other SSE instructions, some of which can be dispatched in pairs of three every clock cycle, this instruction might not always be the optimal choice from a performance perspective.

An example where the usage of the PCLMULQDQ instruction might not lead to a speedup is in the case of polynomial multiplication by x. This can be computed by shifting the input one position to the left (the multiplication by x) and performing a conditional XOR with the reduction polynomial (depending on the bit shifted out). Unfortunately, the SSE instruction set has no bit shift operation shifting the full 128-bit vector. Shifting the two 64-bit, four 32-bit or eight 16-bit in SIMD fashion is possible but the bits shifted out locally are lost. We outline a novel approach (with the SSE instruction in parentheses) to obtain the desired result in the setting of $\mathbb{F}_{2^{128}}$ where we exploit the fact that the second largest exponent of the reduction polynomial is < 32 (which also holds in the setting of $\mathbb{F}_{2^{256}}$). Given an input A we

1. swap the two 64 bit halves of A to t (PSHUFD),
2. create a mask m (either all ones or zeros in each 64-bit half) depending if bits 63 and 127 of t are set (PCMPGTQ),
3. use m to extract the correct 64-bit parts of a precomputed constant $[1, R]$ in t (PAND),
4. shift both 64-bit parts of A left by one bit and store this in s (PSLLQ),
5. perform the actual reduction plus restoring the local carry bit by combining s and t (PXOR).

Here R denotes the hexadecimal representation of the reduction polynomial, excluding the term with the highest exponent, stored in a 64-bit word. Note that this computation might be sped up, depending on the setting, in the following way. Replace step 1 by a byte shuffle (PSHUFD) which moves bits 63 and 127 to bit position 95 and 31 respectively and set the other 14 bytes to zero. The resulting vector, viewed as four 32-bit signed integers, contains two 32-bit words where only the sign bit may be set. Now step 2 can be replaced by using an arithmetic right shift of 31 positions (PSRAD) creating the mask by using the fact that this instruction shifts in the sign bit.

In order to overcome this instruction set limitation (no 128-bit single-bit shift instruction) we tried if polynomial multiplication by x^8 is faster. Now the input needs to be shifted eight bits, which can be performed using a single byte-shuffle instruction. The reduction, a subtraction by $i \cdot R$, where $0 \leq i < 2^8$, depends on the eight bits shifted out. Since the reduction polynomial is constant we can precompute the 256 multiples and use the shifted-out byte as in index for this look-up table. We found that, using our implementation of both approaches, the performance of both polynomial multiplications, by x and x^8, are comparable with a slight advantage when multiplying by x.

2.3 AES-Inspired SHA-3 Candidates

Prompted by recent developments in the cryptanalysis of well-known hash functions MD5 [78] and SHA-1 [77,11], a public competition has been announced by the NIST [55] to develop a new cryptographic hash algorithm intended to replace the current standard SHA-2 [54].

The competition officially started in late 2008 with several submissions from all over the world. As a result, 64 proposals were received, of which 51 met the minimum submission requirements and became the first round candidates. In summer 2009, the number of candidates for the second round was further cut down to a more manageable size of 14 by eliminating the ones having major security or performance flaws. Recently, NIST has announced five finalists, namely BLAKE, GRØSTL, JH, KECCAK and SKEIN; the winner will be announced in 2012.

Among all SHA-3 candidates there exist several algorithms that can potentially benefit from AES-NI [4]: ten out of 51 initial candidates have components which are RIJNDAEL-based and eight out of these algorithms can benefit from AES-NI [4]. Yet, major performance improvement is achieved for only four candidates—ECHO [3], LANE [31], SHAVITE-3 [13] and VORTEX [37]—that use the AES round in its entirety (see [4] for the implementation and performance details). Note that none of these candidates are in the final phase of the competition.

Table 3. The achieved speeds (in cycles per byte) using the AES-NI for the designs considered in this work. Also mentioned are the number of b bytes which are absorbed per compression function call, the primitive employed and how many unique key scheduling calls are made. Predicted speed estimates are based on the results from Table 2. The last column provides additional references.

Algorithm	b	Primitive	Building Block	Key Scheduling	Predicted Speed Range	Achieved Speed	Security Reference
ABREAST-DM [39]	16	Blockcipher	AES-256	two	$11.1 + \epsilon$	11.21	[24,40,44]
DM [51]	32	Blockcipher	RIJNDAEL-256	one	$[6.8, 10.2]$	8.69	[62,16,71,17]
HIROSE-DBL [30]	16	Blockcipher	AES-256	one, shared	9.6	9.82	[30,38]
KNUDSEN–PRENEEL [36]	32	Blockcipher	AES-256	four	10.6	10.58	[58,60]
LANE* (Sec. 3)	64	Permutation	RIJNDAEL-256	fixed	11.7	11.71	[31]
LP231 [66,67]	32	Permutation	RIJNDAEL-256	fixed	$12.6 + \epsilon$	13.04	[66,67,41]
LP362 [66,67]	16	Permutation	AES-128	fixed	$11.8 + \epsilon$	12.09	[66,67,42]
LUFFA* (Sec. 3)	32	Permutation	RIJNDAEL-256	fixed	$8.8 + \epsilon$	10.22	[22]
MDC-2 [19]	16	Blockcipher	AES-128	two	$[9.3, 11.7] + \epsilon$	10.00	[72,35]
MJH [43]	16	Blockcipher	AES-128	one, shared	$6.6 + \epsilon$	7.45	[43]
MJH-DOUBLE [43]	32	Blockcipher	AES-256	one, shared	$4.1 + \epsilon$	4.82	[43]
QPB-DBL [70]	16	Blockcipher	AES-256	one	$9.5 + \epsilon$	14.12	[70]
PEYRIN ET AL.(I) [61]	16	Blockcipher	AES-128	three, shared	$[12.5, 16.3]$	15.09	[68]
PEYRIN ET AL.(II) [61]	32	Blockcipher	AES-256	three, shared	$[7.8, 10.7]$	8.75	[68]
SHRIMPTON–STAM [69]	32	Permutation	RIJNDAEL-256	fixed	12.6	12.39	[69]

3 Implementations of the Target Algorithms

Table 3 contains an overview of the benchmarks we obtained. The measurements have been carried out analogously to [27]; i.e. with the help of the time stamp counter which is read using the `RDTSC` instruction. The presented performance results are an average over thousands of times compressing a random 4KB message. In the sequel, we provide separate treatments for constructions based on a (variable-key) blockcipher versus a permutation (in which case we fix the key of the blockcipher).

Two of the designs considered are based on past SHA-3 candidates. For those, we instantiate the underlying permutation by (fixed-key) RIJNDAEL-256, rather than the originally submitted permutation. For compression functions supporting more than 256-bit output (e.g. KNUDSEN–PRENEEL and LUFFA*) an output transformation (after MD-iteration) can be used to reduce the final output to 256-bit, however we neither implemented nor timed this.

3.1 Blockcipher-Based Constructions

Davies–Meyer (DM). Davies–Meyer (DM) [51] is a single-block-length compression function design. It is one of the most popular ways of creating a secure hash function using a blockcipher: many cryptographic hash functions, including MD5 [64] (for $n = 128$, $k = 512$) and SHA-256 [54] (for $n = 256$, $k = 512$), follow the DM design philosophy.

The first extensive security analysis of DM was performed by Preneel, Govaerts and Vandewalle [62] whose main approach was to attack a class of single-block-length hash/compression functions. The class included DM (for $n = k$), which turned out to successfully resist their efforts. The first security *proof* for DM, on the other hand, was given by Black, Rogaway and Shrimpton [16] (whose technique is later simplified by Stam [71] and jointly published in [17]) who showed that DM indeed enjoys optimal preimage and collision resistance.[1]

DM is one of the most efficient PGV-type compression functions as it allows to run several key schedules independently in the MD-iteration. In our implementations, we exploit this feature; yet we also

[1] We remark that 11 similar compression functions were also proved to be optimally collision and preimage resistant; yet we refrain ourselves from re-implementing all of these variants and focus only on one famous *representative*.

study other possible optimizations. Namely, these are the three flavors of DM that we have considered in our benchmark:

1. Standard iterative approach: compression function calls are made sequentially for each step in the MD-iteration. The compression function evaluation starts with the key schedule and continues with the encryption call. This is illustrated in Fig. 1. Independent key schedule and encryption rounds are interleaved to get more efficient results.
2. Partially pipelined: the encryption call of the current round and the key schedule of the next round are being processed concurrently (see Fig. 2).
3. Fully pipelined: j key schedules are called in parallel for some (integer) $j > 1$ followed by j iterative encryption calls (see Fig. 3). Several experiments were run for varying j and the best result is obtained for $j = 4$. Note that this approach allows to interleave the first encryption round calls with the key scheduling stage to hide latencies and obtain faster results.

Among the three approaches the fully pipelined version gives the best result and is the one reported in Table 3. We included a prediction of the performance of DM based on the vanilla timings of RIJNDAEL-256 provided in Section 2.1. Here the timing for **4K4E** serves as a lower bound, as it makes the encryption calls in parallel. The timing for **4K** plus four times **1E** serves as an upper bound for DM because the first encryption can be scheduled during the four key scheduling stages hiding the instruction dependencies in the encryption improving the overall throughput. (Similar strategies are used for the constructions discussed subsequently. If the predictions in Table 3 include an ϵ, this indicates that certain computations, for instance finite field multiplications, are not considered in the prediction.)

ABREAST-DM. ABREAST-DM and its sister design TANDEM-DM, both proposed in the early 90s [39], are two of the classical examples of double-block-length compression/hash function designs. We only consider ABREAST-DM (see Fig. 8) instantiated with AES-256 for our benchmark. We expect that TANDEM-DM has a slightly worse performance compared to ABREAST-DM due to its sequential structure. In our implementations, we make extensive use of the parallelism inside the ABREAST-DM compression function by calling two key schedules in parallel followed by two concurrent encryption calls (where the 'follow' is on a fine-grained per AES-round basis). Besides this standard approach, there is an opportunity where one creates an asymmetry in the MD-iteration: start with the key schedule for both rows followed by the encryption for either the top or bottom. In all subsequent iterations one can interleave one key schedule and one encryption (from possible different iterations). In our experiments the asymmetric approach did not result in a faster realization over the standard approach. Hence, the prediction for ABREAST-DM is based on the performance numbers for AES-256 in the **2K2E** setting.

HIROSE-DBL. ABREAST-DM suffers from a performance drawback that, although run in parallel, the underlying blockciphers require separate key schedule routines. Hirose's construction [30] overcomes this problem by sharing the key scheduling for the two blockcipher calls (see Fig. 6). Even nicer, the design also enjoys almost optimal collision resistance (recently, it has been proved that the construction has preimage resistance $\Omega(2^{2n})$ [45]). We note that a similar compression function without feed-forward is shown to be almost as collision resistant as HIROSE-DBL [59]. Although we expect a higher efficiency in terms of hardware cost (i.e. area) for the construction without feed-forward, we believe it achieves almost the same speed as HIROSE-DBL in software. In our implementations, we apply the same approach as for ABREAST-DM to HIROSE-DBL (a standard and an asymmetric implementation for the iteration). Again, the standard approach is faster and in accordance with the predicted speed based on the **1K2E** setting for AES-256. Our timings also demonstrate that Hirose's scheme is indeed faster than ABREAST-DM.

MDC-2. MDC-2 [19] is one of the oldest double-block-length hash functions available and it has been specified in the ANSI X9.31 and ISO/IEC 10118-2 standards [1,32]. Although originally designed for use with DES, we consider the obvious generalization (see Fig. 5) where one can use two calls to a single-key blockcipher (where $k = n$ with AES-128).

The compression function of MDC-2 is relatively weak, as collisions and preimages can be found with $\Theta(2^{n/2})$ and $\Theta(2^n)$ queries, respectively. Notwithstanding its early appearance in the literature, the first security proof for MDC-2 (as iterated hash function) is remarkably recent: Steinberger [72] has shown that any collision-finding adversary asking $2^{3n/5-\epsilon}$ (for any $\epsilon > 0$) queries to the underlying blockcipher has a negligible chance of completing a collision. This lower bound then was complemented by the attacks of Knudsen et al. [35], who showed collision and preimage attacks requiring time complexities of $O(2^n/n)$ and $O(2^n)$, respectively. (Reducing the gap between the query complexity lower bound and the time complexity upper bound is still an open problem.)

Since MDC-2 is based on MMO, it is difficult to pipeline multiple MDC-2 compression function calls in the MD-iteration (as we did for DM). Yet, one can benefit from the parallelism naturally present within a single compression function evaluation by making the two blockcipher calls concurrently (corresponding to **2K2E**). This is indeed how we have achieved our best result, matching the predicted speed.

MJH. Recently, an alternative construction called MJH (see Fig. 7) was proposed by Lee and Stam [43]. It is inspired by the compression function of JH [79] (one of the SHA-3 finalists). The main design rationale behind MJH is to reduce the number of key-schedules required in a single compression function evaluation—as in HIROSE-DBL—and call several key schedules in parallel in multiple iterations—as in DM. Obviously, this results in an efficient design. As in the case of MDC-2, the compression function itself does not provide security beyond what a single-length compression function can offer; yet Lee and Stam showed that it enjoys a collision resistance bound of $O(2^{2n/3-\log n})$ in the MD-iteration. Preimages can be found with 2^n queries and in as much time. More interestingly, the security of the construction still holds once the message block (size) to the compression function is doubled (this is what we call the MJH-DOUBLE). This leads to a significantly more efficient scheme, although the cost of key set-up increases. We investigate the performance of MJH in accordance with our optimizations on DM and HIROSE-DBL. Based on our results, we note that MJH-DOUBLE has achieved the best cycle count in our benchmark. We implemented different strategies when interleaving $1 \leq i \leq 8$ iterations of the compression function, the best results are obtained with $i = 2$. Hence, the predictions are based on the setting **2K2E+2E**, ignoring the cost of the polynomial multiplications.

KNUDSEN–PRENEEL. The goal for almost all (multi-block-length) constructions in the 90s has been optimal collision-resistance: a target output size is fixed and the compression function is designed to be collision resistant up to the birthday bound for that digest size. Contrary to this, Knudsen and Preneel [36] adopted a different approach by a priori fixing a particular security target and letting the output size (and number of blockcipher calls) vary in order to guarantee a security level *without* imposing optimal security. To this end, they proposed several constructions with multiple blockcipher calls in parallel using the generator matrices of various linear error correcting codes. Although it was shown [58,60] that the compression functions do not deliver the security level they were designed for, still there exist some constructions satisfying a desirable level of security (when used in wide-pipe mode along with Merkle-Damgård iteration). We consider one of their proposals, which is based on a $[4, 2, 3]$ linear code over $\mathbb{F}_{2^3}$, to show its performance capabilities with AES-NI. We note however that the compression function was not defined explicitly in [36]; we derive the generator matrix based on the works [36,58]. One of the nice features of this construction is that one can call four independent key schedules followed by four independent encryptions where one can interleave the rounds of both operations to hide latencies. This makes it much easier to give an accurate performance estimate since this scenario is exactly the **4K4E** case for AES-256 (cf. Sec. 2.1).

PEYRIN ET AL.-DBL. All the designs considered so far follow a very similar approach: there exist linear pre- and post-processing functions that operate on the blocks of data, interacting with the underlying primitives. The pre-processing function takes the input to the compression function, parses it as blocks and determines (block-wise linearly) the input of the underlying primitives. Similarly, the post-processing function takes the outputs of the underlying primitives, together with the input to the compression func-

tion in case there is a feed-forward, and outputs the digest (that is again based on a linear transformation). Based on this general model, Peyrin et al. [61] determined, under a very general attack-based approach (i.e. only considering time-complexity upper bounds), necessary conditions to have a secure compression function (where they used smaller ideal $2n \to n$ and $3n \to n$ bits compression functions as underlying primitives which are replaced by single-key, resp. double-key, blockciphers in DM mode in our framework). In a later work Peyrin and Seurin [68] followed a more proof-centric approach and derived query-complexity lower and upper bounds for their proposals. They concluded that one needs at least five calls to the (single- or double-key) blockciphers in order to thwart some generic attacks and they proposed some constructions satisfying their criteria. For our purposes, we consider two of their proposals (see Fig. 12) to investigate the performance using AES-NI. In our implementations, we make use of the high parallelism inside a single compression function evaluation by calling several shared key-schedules. In both scenarios the predicted time corresponds to **3K5E**, since among the five encryptions two keys are used twice. This case is not considered in Table 2 and we estimate the performance by considering the performance interval [**3K3E**, **3K3E+2E**] for AES-128 and AES-256 instead.

QPB-DBL. We finish this section with the interesting scenario of constructing a $2n$-bit digest while making only a single call to the blockcipher (theoretically, this would provide optimal efficiency). Lucks [48] provided the first construction of this type, although it is secure only in the iteration (see [59] for a detailed discussion of the security of Lucks' construction). The main practical overhead in Lucks' construction are the costly finite field multiplications that are bound to be performed sequentially. Later, Stam [70] gave another, more practical, construction in the public random function model using a quadratic-polynomial based design (hence the name QPB-DBL). He suggested to use a linear preprocessing function together with a nonlinear post-processing function that uses a quadratic polynomial to evaluate the digest. This construction was generalized [71,46] to the ideal cipher model by replacing the random function by a double-length-key blockcipher running in DM mode (see Fig. 4). Note that we use a slightly different compression function here: in Fig. 4 we assign (V_1, M, V_2) as the input of the compression function whereas (V_2, V_1, M) is suggested in [46]. This change does not violate the security proof, but it has the advantage of allowing increased parallelism in the iteration. As already argued, in the QPB-DBL compression function the main overhead consists of costly finite field multiplications (which we try to minimize by using the features of the new `PCLMULQDQ` instruction). Our tweak allows us to interleave the key-scheduling of round $i+1$ with the two (sequential) finite field multiplications of round i. The predicted performance of QPB-DBL is based on the **1K1E** setting for AES-256 and ignores the relatively high cost of the two finite field multiplications.

3.2 Permutation-Based Constructions

Rogaway and Steinberger's LP and SHRIMPTON–STAM. The Rogaway–Steinberger construction is a class of linearly-determined, permutation-based compression functions $\{0,1\}^{mn} \to \{0,1\}^{rn}$ making k calls to the different[2] permutations π_i for $i \in \{1, \ldots, k\}$ (hence the notation LPmkr throughout). Let (x_i, y_i) denote the input-output pair corresponding to the permutation π_i. The main ingredient of Rogaway and Steinberger's LP design is a $(k+r) \times (k+m)$ matrix A over $\mathbb{F}_{2^n}$ (satisfying an independence criterion [66,67]). This matrix determines the block-wise interaction between the inputs to the compression function (V, M), (x_i, y_i) pairs and the output Z of the compression function in the following way: for the row vector a_i (of A), the inputs to the underlying permutations are determined by the scalar product $x_i = a_i \cdot (V_1, \ldots, V_r, M_1, \ldots, M_{m-r}, y_1, \ldots, y_{i-1})$ whereas the output Z (which is treated as a concatenation of r n-bit blocks Z_i) is computed by $Z_i = a_{k+i} \cdot (V_1, \ldots, V_r, M_1, \ldots, M_{m-r}, y_1, \ldots, y_k)$. Shrimpton and Stam [69] gave a compression function (SS) that can be regarded as an LP231 scheme, although their matrix does not satisfy the independence criterion imposed by Rogaway and Steinberger.

[2] Note that LP framework can be defined for identical permutations as well; yet the corresponding security results are not guaranteed to carry over.

We remark here that the compression function of Shrimpton and Stam also enjoys almost optimal collision resistance (yet suboptimal preimage resistance). In our benchmark, we include two of the LP schemes (LP231 and LP362) as well as SS. For LP231 and SS (see Figs. 13 and 14, respectively), we use the matrices A' and $\tilde{A}$ (both over $\mathbb{F}_{2^{256}}$, see Section 2.2) to define the compression function (the former is the one suggested in [41]). For LP362 we use the matrix A'' (over $\mathbb{F}_{2^{128}}$) given in [42] :

$$A' = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 2 & 1 & 1 & 0 \\ 1 & 1 & 2 & 4 & 2 \end{pmatrix} \quad , \quad \tilde{A} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \end{pmatrix} \quad \text{and} \quad A'' = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 \\ 1 & 2 & 4 & 1 & 2 & 4 & 0 & 1 & 0 \end{pmatrix}.$$

There are multiple ways how one can implement these constructions in practice. We choose to implement LP231 as displayed in Fig. 13, i.e. in three stages where we first run two permutations and a polynomial multiplication by x in parallel, followed by one permutation and polynomial multiplication by x and x^2 and finally the remaining multiplication by x^2. This corresponds to the setting **2E+1E** $+ \epsilon$ for RIJNDAEL-256 on which we base our performance prediction. After experimenting with different strategies we settled on the following regarding LP362. Again three stages are used where we do three, two and one permutation in parallel in every stage. The multiplications are calculated in the last two stages in order to hide the relatively high latencies of especially the single permutation. Hence, the predicted performance is based on the **3E+2E+1E** $+ \epsilon$ setting for AES-128. The implementation of SS is straightforward and is as outlined in Fig. 14. This corresponds to the setting **2E+1E** for RIJNDAEL-256, note that this is the only case where the actual construction (slightly) outperforms the predicted speed This anomaly might be explained by the fact that SS has to load (store) the input (output) only once for both operations while in the performance benchmark setting this has to be done twice.

LANE*. LANE [31] is a permutation-based hash function design (see Fig. 10) submitted to the SHA-3 competition by Indesteege (supported by the COSIC research group). Although some weaknesses have been exploited [50] for the original proposal, the generic compression function of LANE is worth reconsidering due to its capabilities for high parallelism and its suitability for AES-NI. For our purposes, we consider its compression function with 256-bit digest size which is instantiated by eight calls to the fixed-key RIJNDAEL-256 and denoted by LANE*.

For this version, it is not immediate that the known attacks carry over to LANE*. Note that the round function in the original proposal is very similar to the round function of RIJNDAEL-256, yet the overall permutation calls the round function only 6 times (whereas RIJNDAEL-256 is 14 rounds). Nevertheless, the security of the LANE* is known to be sub-optimal: a yield-based adversary can be shown to find collisions and preimages with $O(2^{n/6})$ and $O(2^{n/3})$ queries respectively for an n-bit digest. However, there is still no known algorithm to find collisions and preimages with time complexity less than the generic attacks.

In our implementations, we exploit the high parallelism inside a single compression function evaluation by running several permutation calls in parallel. Although possible, we did not investigate further pipelining options along the MD-iteration due to a sufficient number of independent permutation calls in a single compression function evaluation. The predicted speed for LANE* is based on the setting of **6E+2E** for RIJNDAEL-256. Note that the original version of LANE, performs significantly faster (4.3 cycles per byte) on our platform due to the relatively light permutations given in the submitted version.

LUFFA*. LUFFA [22] is a second round SHA-3 candidate designed by De Cannière and Watanabe. We consider the LUFFA compression function as another permutation-based function that might possibly

benefit from AES-NI once its underlying permutations are modified accordingly (see Fig. 9). To this end, we instantiate the three underlying permutations of LUFFA with fixed-key RIJNDAEL-256 and denote this version by LUFFA*. The security analysis of LUFFA borrows characteristics from the sponge framework [9,2] assuming that the underlying permutations are ideal. We note that the compression function of LUFFA is not ideal and the hash function is claimed to be secure only in the iteration. Moreover, the compression function of LUFFA* outputs 768 bits and one requires an output transformation to reduce the digest to 256-bit.

The implementation of LUFFA* is as outlined in Fig. 9. First the multiplications required in the message injection step are computed (see [22] for a description of how to implement these efficiently), followed by the computation of the three independent permutations. The predicted performance results ($\mathbf{3E} + \epsilon$ using RIJNDAEL-256) is too optimistic, the ϵ incorporates the cost of the multiple polynomial multiplications. Note that our implementation is slightly faster then the original version of LUFFA (which runs at 10.49 cycles per byte) using the fastest implementation (called `SSSE3-PS-2`) submitted to eBASH.

4 Discussion and Conclusion

In this work, we presented the first comprehensive performance comparison of many multi-block-length hash functions (old and new alike) in software on a modern architecture supporting AES-NI. Our results are summarized in Table 3 in conjunction with speed predictions based on the vanilla AES timings from Table 2. Based on these results, we can draw the following conclusions:

1. Our major conclusion is that, when assuming that the underlying primitives behave ideally, one can obtain fast and provably secure blockcipher based hash functions on soon to be mainstream architectures supporting AES-NI. Indeed, the algorithms studied provide reasonable collision and preimage resistance and require between 4 and 15 cycles per byte on our target platform, so in this sense almost *all* of them outperform SHA-256 while several of them are faster than SHA-512.[3] As discussed in the introductions, our results are obtained with the original number of rounds for AES and RIJNDAEL-256. Relative performance results follow by increasing or decreasing the number of rounds, depending on the security margin.
2. Among the blockcipher-based compression functions, DM is the fastest algorithm when optimal security (in terms of proven collision resistance lower bound) is desired. For practical security levels, MJH-DOUBLE significantly outperforms the others (including the permutation-based designs). Note that both constructions require only one key schedule call inside a single compression function evaluation.
3. In the permutation-based setting, the LUFFA* compression function is the fastest, but it is being outperformed by many blockcipher-based constructions. This is partly due to the higher number of primitive calls, but one can argue that our methodology (use AES as is) results in a relatively more conservative security margin for fixed-key constructions. Among the provably secure constructions LP362 performs the best, showing the possibility of achieving higher speed despite the increased number of primitive calls.

Finally, we remark that all the constructions we consider are generic in the sense that they can be instantiated with any secure blockcipher (or permutation, where relevant). Hence, it is well possible that one can achieve better performance with different blockciphers or permutations. In particular, any AES-inspired yet more efficient primitive, for instance a round-reduced version or a tweaked version with more secure and efficient key-scheduling, would result in a faster scheme on our target platform. We believe that our benchmark provides a valuable toolbox to see the relative performance figures for a majority of blockcipher- and permutation-based compression and hash functions.

[3] Compared SHA-256 and SHA-512 speeds (13.90 and 10.47 respectively) are based on the fastest publicly available implementation on eBACS [6] run on Intel Core i5 M 520 (2.4 GHz with AES-NI).

Acknowledgements. This work was supported by the Swiss National Science Foundation under grant numbers 200020-132160, 200021-119776, and 200021-122162 and by the European Commission through the ICT programme under contract ICT-2007-216676 ECRYPT II. We gratefully acknowledge Çağdaş Çalık, for granting us access to the Intel i5 with AES-NI to benchmark our programs and Thorsten Kleinjung for useful discussions on how to optimize the SSE polynomial multiplication by x. We would like to thank the anonymous reviewers for useful comments and suggestions.

References

1. American National Standards Institute. Public key cryptography using reversible algorithms for the financial services industry. American National Standards Institute, 1998.
2. E. Andreeva, B. Mennink, and B. Preneel. Security reductions of the second round SHA-3 candidates. Cryptology ePrint Archive, Report 2010/381, 2010. `http://eprint.iacr.org/`.
3. R. Benadjila, O. Billet, H. Gilbert, G. Macario-Rat, T. Peyrin, M. Robshaw, and Y. Seurin. SHA-3 proposal: ECHO, 2009.
4. R. Benadjila, O. Billet, S. Gueron, and M. J. B. Robshaw. The Intel AES instructions set and the SHA-3 candidates. In *Asiacrypt 2009*, volume 5912 of *LNCS*, pages 162–178. Springer, 2009.
5. D. J. Bernstein. Cache-timing attacks on AES, 2005. `http://cr.yp.to/papers.html#cachetiming`.
6. D. J. Bernstein and T. Lange, (editors). eBACS: ECRYPT Benchmarking of Cryptographic Systems. `http://bench.cr.yp.to`, 2010.
7. D. J. Bernstein and P. Schwabe. New AES software speed records. In *Indocrypt 2008*, volume 5365 of *LNCS*, pages 322–336. Springer, 2008.
8. G. Bertoni, L. Breveglieri, R. Farina, and F. Regazzoni. Speeding up AES by extending a 32 bit processor instruction set. In *Application-specific Systems, Architectures and Processors*, pages 275–282. IEEE Computer Society, 2006.
9. G. Bertoni, J. Daemen, M. Peeters, and G. Van Assche. On the indifferentiability of the sponge construction. In *Eurocrypt 2008*, volume 4965 of *LNCS*, pages 181–197. Springer, 2008.
10. E. Biham. A fast new DES implementation in software. In *FSE 1997*, volume 1267 of *LNCS*, pages 260–272, 1997.
11. E. Biham, R. Chen, A. Joux, P. Carribault, C. Lemuet, and W. Jalby. Collisions of SHA-0 and reduced SHA-1. In *Eurocrypt 2005*, volume 3494 of *LNCS*, pages 36–57. Springer, 2005.
12. E. Biham and O. Dunkelman. A framework for iterative hash functions – HAIFA. Presented at *Second NIST Cryptographic Hash Workshop, August 24–25, 2006, Santa Barbara, California, USA*.
13. E. Biham and O. Dunkelman. The SHAvite-3 hash function, 2009.
14. A. Biryukov and D. Khovratovich. Related-key cryptanalysis of the full AES-192 and AES-256. In *Asiacrypt 2009*, volume 5912 of *LNCS*, pages 1–18. Springer, 2009.
15. A. Biryukov, D. Khovratovich, and I. Nikolic. Distinguisher and related-key attack on the full AES-256. In *Crypto 2009*, volume 5677 of *LNCS*, pages 231–249, 2009.
16. J. Black, P. Rogaway, and T. Shrimpton. Black-box analysis of the block-cipher-based hash-function constructions from PGV. In *Crypto 2002*, volume 2442 of *LNCS*, pages 320–335. Springer, 2002.
17. J. Black, P. Rogaway, T. Shrimpton, and M. Stam. An analysis of the blockcipher-based hash functions from PGV. *Journal of Cryptology*, 23(4):519–545, 2010.
18. A. Bogdanov, G. Leander, C. Paar, A. Poschmann, M. J. B. Robshaw, and Y. Seurin. Hash functions and RFID tags: Mind the gap. In E. Oswald and P. Rohatgi, editors, *CHES '08*, volume 5154 of *LNCS*, pages 283–299. Springer, 2008.
19. B. Brachtl, D. Coppersmith, M. Hyden, S. Matyas, Jr., C. Meyer, J. Oseas, S. Pilpel, and M. Schilling. Data authentication using modification detection codes based on a public one-way encryption function. U.S. Patent No 4,908,861, 1990.
20. J. Daemen and V. Rijmen. *The design of Rijndael*. Springer, 2002.
21. I. Damgård. A design principle for hash functions. In *Crypto 1989*, volume 435 of *LNCS*, pages 416–427. Springer, 1990.
22. C. De Cannière, H. Sato, and D. Watanabe. Hash function Luffa: Supporting document. Submission to NIST (Round 2), 2009.
23. Ecrypt II yearly report on algorithms and key sizes. Ecrypt II European Network of Excellence in Cryptology II, 30 March 2010.
24. E. Fleischmann, M. Gorski, and S. Lucks. Security of cyclic double block length hash functions. In *Cryptography and Coding 2009*, volume 5921 of *LNCS*, pages 153–175. Springer, 2009.
25. A. Fog. Instruction tables, lists of instruction latencies, throughputs and microoperation breakdowns for Intel, AMD and VIA CPUs. `http://www.agner.org/optimize/`, 2010.
26. S. Gueron. Intel's new AES instructions for enhanced performance and security. In *FSE 2009*, volume 5665 of *LNCS*, pages 51–66. Springer, 2009.
27. S. Gueron. Intel advanced encryption standard (AES) instructions set. Technical report, Intel, 2010. `http://software.intel.com/en-us/articles/intel-advanced-encryption-standard-aes-instructions-set/`.

28. S. Gueron and M. E. Kounavis. Intel carry-less multiplication instruction and its usage for computing the GCM mode. Technical report, Intel, 2010. http://software.intel.com/en-us/articles/intel-carry-less-multiplication-instruction-and-its-usage-for-computing-the-gcm-mode/.
29. M. Hamburg. Accelerating AES with vector permute instructions. In *CHES*, volume 5747 of *LNCS*, pages 18–32. Springer, 2009.
30. S. Hirose. Some plausible constructions of double-block-length hash functions. In *FSE 2006*, volume 4047 of *LNCS*, pages 210–225. Springer, 2006.
31. S. Indesteege. The LANE hash function. Submission to NIST, 2008.
32. International Organization for Standardization. ISO/IEC 10118-2:2000.information technology-security techniques-hash functions-hash functions using an n-bit block cipher. International Organization for Standardization, Geneva, Switzerland, 2000, First released in 1992.
33. E. Käsper and P. Schwabe. Faster and timing-attack resistant AES-GCM. In *CHES 2009*, volume 5747 of *LNCS*, pages 1–17, 2009.
34. D. Khovratovich. *New Approaches to the Cryptanalysis of Symmetric Primitives*. PhD thesis, University of Luxembourg, 2010.
35. L. R. Knudsen, F. Mendel, C. Rechberger, and S. S. Thomsen. Cryptanalysis of MDC-2. In *Eurocrypt 2009*, volume 5479 of *LNCS*, pages 106–120. Springer, 2009.
36. L. R. Knudsen and B. Preneel. Construction of secure and fast hash functions using nonbinary error-correcting codes. *IEEE Transactions on Information Theory*, 48(9):2524–2539, 2002.
37. M. Kounavis and S. Gueron. Vortex: A new family of one way hash functions based on rijndael rounds and carry-less multiplication. Submission to NIST, 2008.
38. M. Krause, F. Armknecht, and E. Fleischmann. Preimage resistance beyond the birthday barrier – the case of blockcipher based hashing. Cryptology ePrint Archive, Report 2010/519, 2010. http://eprint.iacr.org/.
39. X. Lai and J. L. Massey. Hash functions based on block ciphers. In *Eurocrypt '92*, volume 658 of *LNCS*, pages 55–70. Springer, 1993.
40. J. Lee and D. Kwon. The security of Abreast-DM in the ideal cipher model. Cryptology ePrint Archive, Report 2009/225, 2009. http://eprint.iacr.org/.
41. J. Lee and J. H. Park. Adaptive preimage resistance and permutation-based hash functions. Cryptology ePrint Archive, Report 2009/066, 2009. http://eprint.iacr.org/.
42. J. Lee and J. H. Park. Preimage resistance of LPmkr with $r = m - 1$. *Information Processing Letters*, 110(14-15):602–608, 2010.
43. J. Lee and M. Stam. MJH: a faster alternative to MDC-2. In *CT-RSA '11*, volume 6558 of *LNCS*, pages 213–236. Springer, 2011.
44. J. Lee, M. Stam, and J. Steinberger. The collision security of Tandem-DM in the ideal cipher model. Cryptology ePrint Archive, Report 2010/409, 2010. http://eprint.iacr.org/.
45. J. Lee, M. Stam, and J. Steinberger. The preimage security of double-block-length compression functions. Cryptology ePrint Archive, Report 2011/224, 2011. http://eprint.iacr.org/.
46. J. Lee and J. P. Steinberger. Multi-property-preserving domain extension using polynomial-based modes of operation. In *Eurocrypt 2010*, volume 6110 of *LNCS*, pages 573–596. Springer, 2010.
47. S. Lucks. A failure-friendly design principle for hash functions. In *ASIACRYPT '05*, volume 3788 of *LNCS*, pages 474–494. Springer, 2005.
48. S. Lucks. A collision-resistant rate-1 double-block-length hash function. In E. Biham, H. Handschuh, S. Lucks, and V. Rijmen, editors, *Symmetric Cryptography*, number 07021 in Dagstuhl Seminar Proceedings, Dagstuhl, Germany, 2007. Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl, Germany.
49. R. Manley, P. Magrath, and D. Gregg. Code generation for hardware accelerated AES. In *Application-specific Systems Architectures and Processors (ASAP), 21st IEEE International Conference on*, pages 345–348, 2010.
50. K. Matusiewicz, M. Naya-Plasencia, I. Nikolic, Y. Sasaki, and M. Schläffer. Rebound attack on the full Lane compression function. In *Asiacrypt 2009*, volume 5912 of *LNCS*, pages 106–125. Springer, 2009.
51. A. Menezes, P. van Oorschot, and S. Vanstone. *CRC-Handbook of Applied Cryptography*. CRC Press, 1996.
52. R. C. Merkle. One way hash functions and DES. In *CRYPTO 1989*, volume 435 of *LNCS*, pages 428–446. Springer, 1990.
53. NIST. FIPS-197: Advanced encryption standard (AES), 2001. http://www.csrc.nist.gov/publications/fips/fips197/fips-197.pdf.
54. NIST. Secure hash standard. FIPS 180-2, NIST, http://www.itl.nist.gov/fipspubs/fip180-2.htm, August 2002.
55. NIST. Cryptographic hash algorithm competition. http://csrc.nist.gov/groups/ST/hash/sha-3/index.html, 2008.
56. D. A. Osvik, J. W. Bos, D. Stefan, and D. Canright. Fast software AES encryption. In *FSE 2010*, volume 6147 of *LNCS*, pages 75–93. Springer, 2010.
57. D. A. Osvik, A. Shamir, and E. Tromer. Cache attacks and countermeasures: The case of AES. In *CT-RSA 2006*, volume 3860 of *LNCS*, pages 1–20. Springer, 2006.
58. O. Özen, T. Shrimpton, and M. Stam. Attacking the Knudsen-Preneel compression functions. In *FSE 2010*, volume 6147 of *LNCS*, pages 94–115. Springer, 2010.

59. O. Özen and M. Stam. Another glance at double-length hashing. In *Cryptography and Coding 2009*, volume 5921 of *LNCS*, pages 176–201. Springer, 2009.
60. O. Özen and M. Stam. Collision attacks against the Knudsen-Preneel compression functions. In *To appear in Asiacrypt 2010*, LNCS. Springer, 2010.
61. T. Peyrin, H. Gilbert, F. Muller, and M. J. B. Robshaw. Combining compression functions and block cipher-based hash functions. In *Asiacrypt 2006*, volume 4284 of *LNCS*, pages 315–331. Springer, 2006.
62. B. Preneel, R. Govaerts, and J. Vandewalle. Hash functions based on block ciphers: A synthetic approach. In *Crypto '93*, volume 773 of *LNCS*, pages 368–378. Springer, 1994.
63. M. O. Rabin. Digitalized signatures. In *Foundations of Secure Computations*, pages 155–166. Academic Press, 1978.
64. R. Rivest. The MD5 message-digest algorithm, request for comments (RFC) 1320. Technical report, Internet Activities Board, Internet Privacy Task Force, 1992.
65. P. Rogaway and T. Shrimpton. Cryptographic hash-function basics: Definitions, implications, and separations for preimage resistance, second-preimage resistance, and collision resistance. In B. K. Roy and W. Meier, editors, *FSE 2004*, volume 3017 of *LNCS*, pages 371–388. Springer, 2004.
66. P. Rogaway and J. Steinberger. Security/efficiency tradeoffs for permutation-based hashing. In *Eurocrypt 2008*, volume 4965 of *LNCS*, pages 220–236. Springer, 2008.
67. P. Rogaway and J. P. Steinberger. Constructing cryptographic hash functions from fixed-key blockciphers. In *Crypto 2008*, volume 5157 of *LNCS*, pages 433–450. Springer, 2008.
68. Y. Seurin and T. Peyrin. Security analysis of constructions combining FIL random oracles. In *FSE 2007*, volume 4593 of *LNCS*, pages 119–136. Springer, 2007.
69. T. Shrimpton and M. Stam. Building a collision-resistant compression function from non-compressing primitives. In *International Colloquium on Automata, Languages and Programming 2008*, volume 5126 of *LNCS*, pages 643–654. Springer, 2008.
70. M. Stam. Beyond uniformity: Better security/efficiency tradeoffs for compression functions. In *Crypto 2008*, volume 5157 of *LNCS*, pages 397–412. Springer, 2008.
71. M. Stam. Blockcipher-based hashing revisited. In *FSE 2009*, volume 5665 of *LNCS*, pages 67–83. Springer, 2009.
72. J. P. Steinberger. The collision intractability of MDC-2 in the ideal-cipher model. In *Eurocrypt 2007*, volume 4515 of *LNCS*, pages 34–51. Springer, 2007.
73. S. Tillich and J. Großschädl. Instruction set extensions for efficient AES implementation on 32-bit processors. In *CHES 2006*, volume 4249 of *LNCS*, pages 270–284. Springer, 2006.
74. S. Tillich, J. Großschädl, and A. Szekely. An instruction set extension for fast and memory-efficient AES implementation. In *Communications and Multimedia Security*, volume 3677 of *LNCS*, pages 11–21. Springer, 2005.
75. S. Tillich and C. Herbst. Boosting AES performance on a tiny processor core. In *CT-RSA*, volume 4964 of *LNCS*, pages 170–186, 2008.
76. E. Tromer, D. A. Osvik, and A. Shamir. Efficient cache attacks on AES, and countermeasures. *Journal of Cryptology*, 23:37–71, 2010.
77. X. Wang, Y. L. Yin, and H. Yu. Finding collisions in the full SHA-1. In *Crypto 2005*, volume 3621 of *LNCS*, pages 17–36. Springer, 2005.
78. X. Wang and H. Yu. How to break MD5 and other hash functions. In *Eurocrypt 2005*, volume 3494 of *LNCS*, pages 19–35. Springer, 2005.
79. H. Wu. The hash function JH. Submission to NIST (updated), 2009.

A Security Considerations

There exist multiple security notions for a cryptographic hash function to satisfy; we only consider the collision-and preimage-resistance and the relevant adversarial models. A *preimage-finding adversary* is an algorithm with access to one or more oracles and whose goal is to find a preimage of some specified compression/hash function output.[4] Similarly, a *collision-finding adversary* is an algorithm whose goal is to find collisions in some specified compression or hash function.

To get a better picture, we consider adversaries in two scenarios: the information- and the complexity-theoretic. For the former, the only resource of interest is the number of queries made to their oracles, the so called query-complexity, where the adversaries are considered computationally unbounded. For the latter, on the other hand, we consider the actual runtime of the adversarial algorithm (with respect to a reasonable computational model). By convention, we assume that for optimally secure constructions the only valid attacks are the generic attacks which require $\Theta(2^{s/2})$ and $\Theta(2^s)$ queries/time to

[4] Note that there exist several definitions of preimage resistance formalized in [65].

Table 4. The comparison of all the relevant multi-block length designs, the SHA-3 finalists, SHA-256 and SHA-512. Security claims are given in accordance with the conventions introduced in Appendix A. The relevant entries illustrate the bit security. Briefly, lower bounds correspond to the proven query-complexity lower bounds (under the assumption that the underlying primitives are ideal) and query/time-complexity upper bounds corresponding to the best known attacks matching these bounds. The performance numbers for SHA-256, SHA-512 and SHA-3 finalists (with at least 256-bit digest) are obtained from the submitted codes to eBACS [6] run on Intel Core i5 M 520 (2.4 GHz with AES-NI). Note that for SHA-3 finalists we simply refer to the fastest result on the same platform without differentiating the different digest-size versions. The first column shows the corresponding algorithm. The number of b bytes which are absorbed per compression function call is shown together with the primitive employed, the achieved speed using the AES instructions (with or without fixed-key) and how many unique key scheduling calls are made. The remainder of the table shows the known results in terms of the collision and the preimage resistance.

Algorithm	b	**Building Block**	**Key Scheduling**	**Speed**	**Collision Resistance**						**Preimage Resistance**					
					Compression Function			Mode of Operation			Compression Function			Mode of Operation		
					Query		Time	Query		Time	Query		Time	Query		Time
				Cycles per byte	Lower Bound	Upper Bound		Lower Bound	Upper Bound		Lower Bound	Upper Bound		Lower Bound	Upper Bound	
SHA-256	64	blockcipher	one	13.90	127	128	128	127	128	128	255	256	256	255	256	256
SHA-512	128	blockcipher	one	10.47	255	256	256	255	256	256	511	512	512	511	512	512
BLAKE-256	64	blockcipher	one	5.93	127	128	128	127	128	128	255	256	256	255	256	256
GRØSTL-256	64	permutation	fixed	17.80	127	128	170	127	128	128	255	256	256	255	256	256
JH-256	64	permutation	fixed	13.41	-	1	1	128	128	128	-	0	0	170	256	256
KECCAK-C512	64	permutation	fixed	10.09	-	1	1	127	128	128	-	0	0	255	256	256
SKEIN-512	64	blockcipher	one	5.07	255	256	256	255	256	256	511	512	512	511	512	512
ABREAST-DM	16	AES-256	two	11.21	125	128	128	125	128	128	127	256	256	127	256	256
DM	32	RIJNDAEL-256	one	8.69	127	128	128	127	128	128	255	256	256	255	256	256
HIROSE-DBL	16	AES-256	one, shared	9.82	126	128	128	126	128	128	171	256	256	172	256	256
KNUDSEN–PRENEEL	32	AES-256	four	10.58	-	128	128	-	128	128	-	256	256	255	256	256
LANE*	64	RIJNDAEL-256	fixed	11.71	-	43	128	-	128	128	-	86	256	-	256	256
LP231	32	RIJNDAEL-256	fixed	13.04	127	128	128	127	128	128	172	172	256	172	181	256
LP362	16	AES-128	fixed	12.09	81	84	128	81	88	128	102	106	256	102	116	256
LUFFA*	32	RIJNDAEL-256	fixed	10.22	-	1	1	-	128	128	-	0	0	-	256	256
MDC-2	16	AES-128	two	10.00	-	64	64	77	120	120	-	128	128	-	128	128
MJH	16	AES-128	one, shared	7.45	-	64	64	80	128	128	-	128	128	-	128	128
MJH-DOUBLE	32	AES-256	one, shared	4.82	-	64	64	80	128	128	-	128	128	-	128	128
QPB-DBL	16	AES-256	one	14.12	120	128	128	120	128	128	-	256	256	-	256	256
PEYRIN ET AL.(I)	16	AES-128	three, shared	15.09	86	86	128	86	128	128	172	172	256	172	172	256
PEYRIN ET AL.(II)	32	AES-256	three, shared	8.75	86	96	128	86	128	128	172	172	256	172	172	256
SS	32	RIJNDAEL-256	fixed	12.39	127	128	128	127	128	128	127	172	256	127	181	256

find collisions and preimages for a hash/compression function of s-bit digest respectively. From an information-theoretic point of view, the compression functions we consider are not always optimal. Yet, in the complexity-theoretic setting, almost all of them are considered to be optimally secure in the sense that there exist no known algorithm to find collisions and preimages faster than the generic methods.

In the literature, known security results contain both models. Most of the security lower bounds—for either security notions—are given in the information-theoretic setting whereas the upper bounds are obtained by concrete *attacks* either in the information- or complexity-theoretic model. The designs considered in this work provide at least 75-bit security in the information-theoretic setting for collision resistance (the security margin increases up to 120-bit for the case of preimage resistance). For some of the constructions that is the best one can (currently) prove, but for most of the constructions the best practical attack (in the complexity-theoretic setting) is in fact a generic birthday attack (even when instantiated with AES). We summarize the known security results (for 256-bit digest) for our target algorithms in Table 4 along with the currently available results against SHA-256, SHA-512 and SHA-3 finalists.

Note that for the security of the mode of operation, we assume that the Merkle–Damgård iteration is taking place for most of the compression function designs and the respective security preservations [52,21,65] hold. For the designs proposed along with different mode of operations, we simply assume that the original mode is taking place and the corresponding security reductions follow [9,12].

B Illustrations of Related Compression Functions

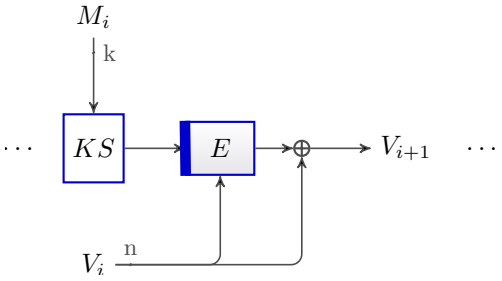

Fig. 1. The DM compression function, $n = k = 256$.The underlying blockcipher is RIJNDAEL-256. KS shows the key scheduling algorithm; KS and E are called iteratively, they are interleaved in round function level.

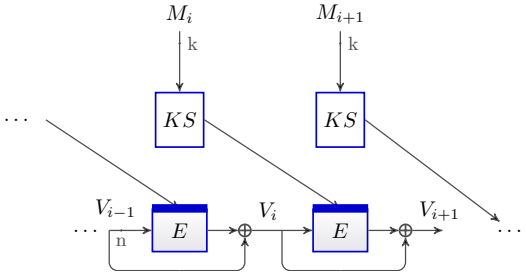

Fig. 2. The DM compression function, with $n = k = 256$, is illustrated inside MD-iteration. The underlying blockcipher is RIJNDAEL-256. KS shows the key scheduling algorithm; here the key schedule for the next round is computed concurrently with the blockcipher call of the current round. Note that the two are completely independent.

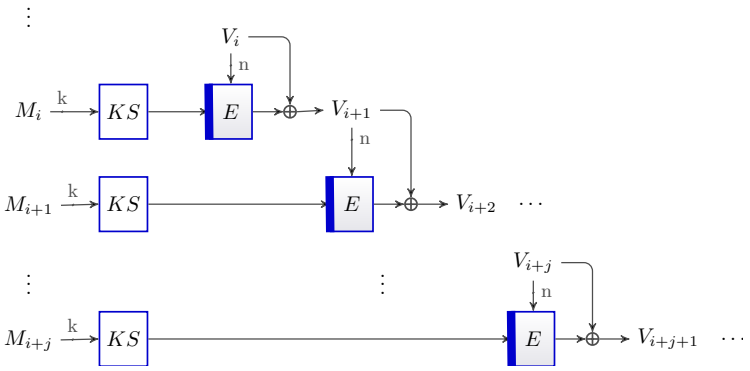

Fig. 3. The DM compression function, with $n = k = 256$, is illustrated inside MD-iteration (for $j + 1$ iterations). The underlying blockcipher is RIJNDAEL-256. Here KS shows the key scheduling algorithm and the $j + 1$ of those are running in parallel; whereas the encryption calls are made sequentially.

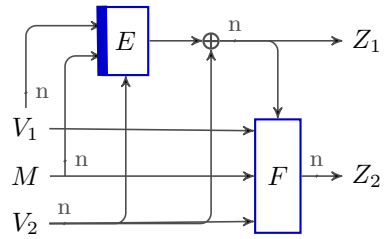

Fig. 4. The QPB-DBL compression function, $n = 128$. The underlying blockcipher is AES-256 and $Z_2 = Z_1(V_1Z_1 + M) + V_2$.

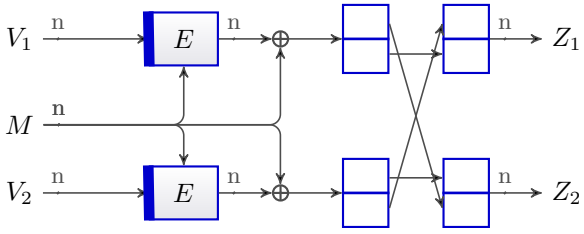

Fig. 5. The MDC-2 compression function, $n = 128$. The underlying blockcipher is AES-128.

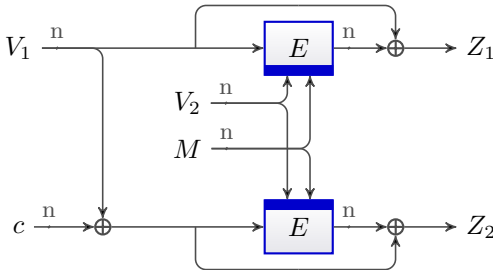

Fig. 6. The HIROSE-DBL compression function, $n = 128$. The underlying blockcipher is AES-256 and $c \in \{0,1\}^n \setminus \{0\}^n$.

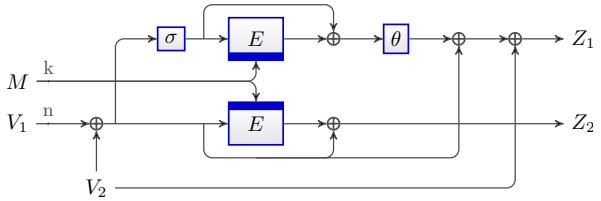

Fig. 7. The MJH and MJH-DOUBLE compression functions for $n = 128$. In the former, $k = n = 128$ whereas in the latter $k = 2n = 256$. The underlying blockciphers are AES-128 and AES-256 respectively, σ is an involution (e.g. an addition with a non-zero constant) and θ is a multiplication with a constant $c \in \mathbb{F}_{2^{128}}$ and $c \notin \mathbb{F}_2$.

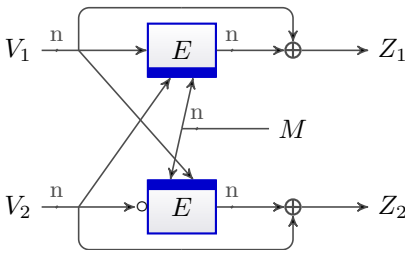

Fig. 8. The ABREAST-DM compression function, $n = 128$. The underlying blockcipher is AES-256 and $\circ$ denotes the bitwise complementation.

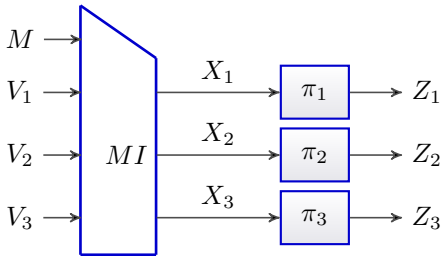

Fig. 9. The LUFFA-256 compression function, the horizontal lines carry 256 bits. The underlying permutations are obtained from the fixed-key RIJNDAEL-256 with varying keys. The message injection step MI is defined as follows:
$X_i = V_i \oplus (0\text{x}02 \cdot (V_1 \oplus V_2 \oplus V_3)) \oplus 0\text{x}02^i \cdot M$, for $i \leq 1 \leq 3$. Note that the hash function outputs 256 bits by performing an output transformation. We refer to [22] for all details, especially how the multiplication by $0\text{x}02$ is defined.

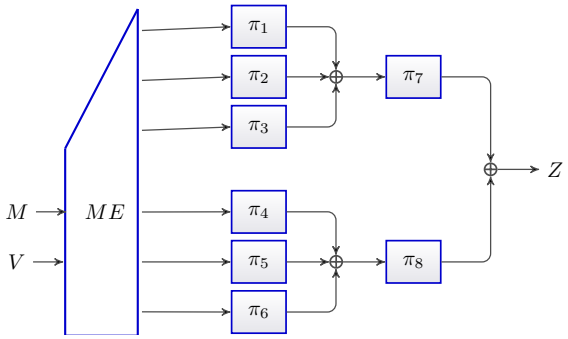

Fig. 10. The LANE-256 compression function with $M \in \{0,1\}^{512}$ and $V, Z \in \{0,1\}^{256}$. Horizontal lines carry 256 bits. The underlying permutations are obtained from the fixed-key RIJNDAEL-256 with varying keys. Here ME denotes the so called message expansion algorithm.

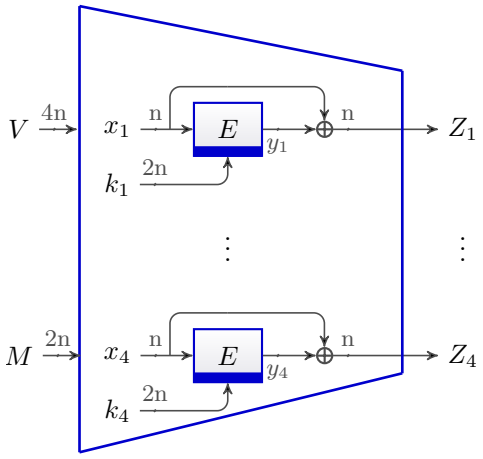

Fig. 11. The KNUDSEN–PRENEEL compression function, $n = 128$. The underlying blockcipher is AES-256. For $V = V_1||\ldots||V_4 \in \{0,1\}^{512}$, $M = M_1||M_2 \in \{0,1\}^{256}$ and different constants $c_i \in \{0,1\}^{128}$, we define:
$(x_1, k_1) = (V_1 \oplus c_1, V_2||V_3)$,
$(x_2, k_2) = (V_4 \oplus c_2, M_1||M_2)$,
$(x_3, k_3) = (V_3 \oplus M_1 \oplus c_3, V_1 \oplus V_2 \oplus M_2||V_2 \oplus V_3 \oplus V_4 \oplus M_1)$,
$(x_4, k_4) = (V_1 \oplus V_3 \oplus M_1 \oplus M_2 \oplus c_4, V_1 \oplus V_4 \oplus M_1 \oplus M_2||V_2 \oplus V_4 \oplus M_2)$.

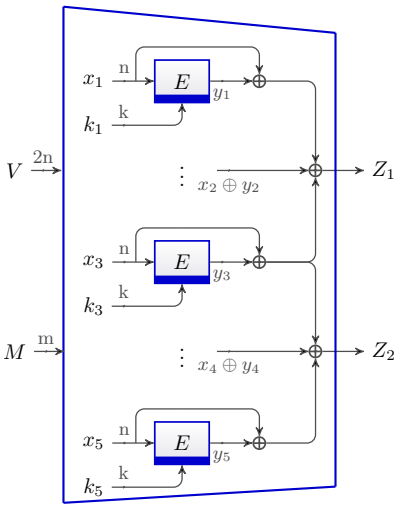

Fig. 12. The PEYRIN ET AL. compression function $H : \{0,1\}^{2n+m} \to \{0,1\}^{2n}$. The underlying five blockciphers operate on n bits and they support k-bit keys each. We consider two scenarios (for different constants $c_i \in \{0,1\}^{128}$):
(i) $m = k = n = 128$ (using AES-128), for $V = V_1||V_2$ set
$(x_1, k_1) = (V_1 \oplus c_1, V_2)$, $(x_2, k_2) = (V_2 \oplus c_2, M)$,
$(x_3, k_3) = (M \oplus c_3, V_1 \oplus V_2)$, $(x_4, k_4) = (V_1 \oplus c_4, M)$,
$(x_5, k_5) = (V_1 \oplus c_5, V_2)$.
(ii) $m = k = 2n = 256$ (using AES-256), for $V = V_1||V_2$, $M = M_1||M_2$ set
$(x_1, k_1) = (V_1 \oplus c_1, V_2||M_1)$, $(x_2, k_2) = (V_1 \oplus c_2, V_2||M_2)$,
$(x_3, k_3) = (V_1 \oplus c_3, M_1||M_2)$, $(x_4, k_4) = (V_1 \oplus c_4, V_2||M_1)$,
$(x_5, k_5) = (V_2 \oplus c_5, M_1||M_2)$.

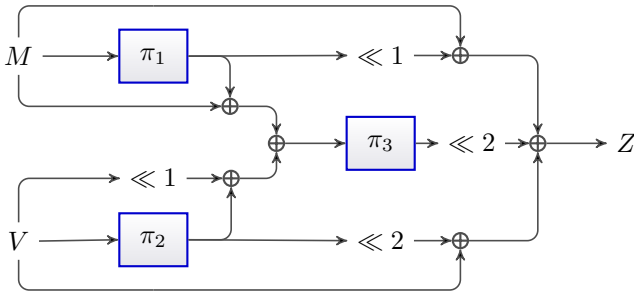

Fig. 13. The ROGAWAY–STEINBERGER's LP231 compression function is illustrated, all lines carry 256 bits. The underlying fixed-key permutations π_1, π_2, π_3 are derived from fixed-key RIJNDAEL-256. Here $\ll 1$ and $\ll 2$ denote the polynomial multiplication with x and x^2 respectively.

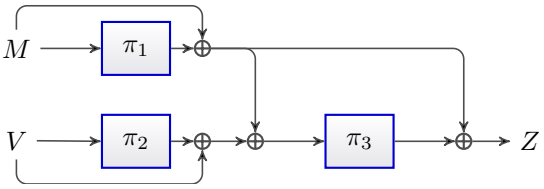

Fig. 14. The SHRIMPTON–STAM compression function is illustrated, all lines carry 256 bits. The underlying fixed-key permutations π_1, π_2, π_3 are derived from fixed-key RIJNDAEL-256.