The $m = n + 1$ Boundary of EME: A Splicing Distinguisher for the Unrefreshed EME-Core Extension and Its Linear-Map Generalization

Jiadong Han[1], Peng Wang[1]

School of Cryptology, University of Chinese Academy of Sciences, Beijing, China
hanjiadong25@mails.ucas.ac.cn, p-wang@ucas.ac.cn

Abstract. EME is a parallelizable encrypt–mix–encrypt wide-block construction proved secure for m-block messages only in the range $m \leq n$, where n is the block length of the underlying block cipher. Halevi and Rogaway justified this restriction by giving a splicing distinguisher for $m \geq n + 2$, but left open the first excluded length, $m = n + 1$. We resolve that boundary for the direct, *unrefreshed* EME-core extension beyond its specified $m \leq n$ domain: the original formulas are applied to $m = n + 1$ blocks while continuing to use the same global mask. Under one fixed tweak, two encryption queries and one decryption query distinguish this extension from a random permutation. The result applies to this unrefreshed extension; refreshed variants such as EME* and IEEE EME2-AES are outside the scope of the distinguisher. The boundary attack is not a shortened form of the known zero-sum attack. At $m = n+1$, the non-first coefficients have no nonempty zero-sum; instead, they form a basis and therefore uniquely represent the coefficient 1. Combining this representation with the special equation for the first mixed block gives the cancellation needed for a ciphertext splice. The same mechanism applies to Sarkar's EMME framework: when multiplication by the field element x is replaced by a linear operator ψ with degree-n minimal polynomial, the minimal-polynomial relation supplies the corresponding operator identity. EME* and IEEE EME2-AES avoid this setting by refreshing the mask at chunk boundaries. The attack identifies the algebraic obstruction that such refresh steps avoid.

Keywords: Wide-block encryption · tweakable SPRP · EME · EME* · EME2-AES · linear recurrences · splicing attacks

1 Introduction

Length-preserving encryption is a natural requirement in storage and legacy protocols: for example, a disk sector must be encrypted without changing its size, while its sector number can serve as a public tweak so that different sectors are processed under separate permutations. Since a deterministic length-preserving scheme cannot provide semantic security in the usual randomized-encryption

sense, the conventional goal is a tweakable strong pseudorandom permutation on each supported message length, leading to *tweakable enciphering schemes* (TESs). Over the last two decades, representative block-cipher-based TESs have included the encrypt–mix–encrypt designs CMC [6], EME [7], and EME* [8], as well as hash–counter–hash designs such as XCB [11] and HCTR [14]. EME was designed in particular for disk-sector encryption, and EME* extended it to almost arbitrary lengths; HCTR2 [3] later repaired specification and proof problems in HCTR while improving its flexibility and performance.

More recently, NIST adopted the term *accordion mode* [2] for a block-cipher-based, tweakable variable-input-length strong pseudorandom permutation. Its accordion initiative and 2024 workshop stimulated further designs, including the docked-double-decker instances ddd-AES [5], the tweakable-block-cipher-based HCTR+ [4], and the double-block HCTR variant DbHCTR [10,13].

EME and the encrypt–mix–encrypt paradigm. The feature of EME central to this paper is its parallel encrypt–mix–encrypt structure: two ECB passes surround a lightweight nonlinear mixing layer, with the mode specified on mn-bit strings for $m \in [1..n]$. Its position masks are generated from the coefficient elements

$$1, \alpha, \alpha^2, \ldots, \alpha^{n-1}$$

which are linearly independent over $\mathbb{F}_2$. Here α is the field element represented by x, equivalently the multiplication-by-x map on $\mathbb{F}_{2^n} = \mathbb{F}_2[x]/(p(x))$ for an irreducible degree-n polynomial $p(x)$. This linear independence underlies the specified bound $m \leq n$. Sarkar subsequently abstracted this mechanism in a general ECB–mix–ECB framework [12], called EMME here, by replacing multiplication by powers of α with iterates of a linear map ψ. When ψ has a primitive degree-n minimal polynomial, its first n iterates retain the nonzero-sum property needed by the EME proof [12]. Thus EME is the field-multiplication instance, whereas EMME exposes the linear-map condition used later in our generalization.

Halevi and Rogaway [7] also showed that the direct, unrefreshed continuation of EME is distinguishable for $m \geq n + 2$, using a nonempty zero-sum relation among coefficients of the non-first blocks, but explicitly left the borderline case $m = n + 1$ open.

Relation to standardized variants. Three related objects should be kept separate. Original EME is specified only for $m \leq n$. The target of this paper is the formal unrefreshed extension at the boundary, in the sense of Section 2: the original EME core formulas are applied to $m = n + 1$ while the same global mask is carried into the extra block. EME* [8] and IEEE EME2-AES [9] are different arbitrary-length designs: they cross chunk boundaries by recomputing masks, so they do not contain the old-mask continuation used by the splice.

In the terminology of NIST IR 8552, EME, EME*, and EME2-AES are relevant examples in the development of accordion mode [2]. IR 8552 specifies requirements rather than an algorithm; the standardized construction closest to the present discussion is IEEE 1619.2-2025 EME2-AES [9], a wide-block encryption mode with associated data for shared storage media. These developments do

not change the scope of our result, which concerns only the direct, unrefreshed EME-core extension.

The $m = n+1$ boundary. We revisit this boundary. At $m = n+1$, the coefficients

$$\alpha, \alpha^2, \ldots, \alpha^n$$

do not contain a nonempty zero-sum relation; indeed, they form a basis of $\mathbb{F}_{2^n}$ over $\mathbb{F}_2$. The original $m = n + 2$ argument therefore does not apply directly. Because these coefficients form a basis, however, the field element 1 can be represented as a sum of some of them. If

$$1 = \bigoplus_{k \in S} \alpha^{k-1} \qquad \text{for some } S \subseteq \{2, \ldots, n+1\},$$

then the special first-block mixing equation supplies the missing 1-coefficient term. This yields the cancellation needed for a ciphertext splice and, ultimately, a distinguisher at the first length outside the specified EME domain. In EMME, the same basis argument becomes an identity derived from the minimal polynomial of ψ.

Contributions. The main contributions are as follows.

1. *A distinguisher at $m = n + 1$.* The direct, unrefreshed EME-core extension is distinguishable at the first length outside the original proof range. The distinguisher uses two encryption queries and one decryption query under a fixed tweak.
2. *The algebraic source of the boundary attack.* We explain why the unrefreshed boundary attack is not the known $m \geq n+2$ zero-sum attack in shorter form. Instead, it relies on the representation of the coefficient 1 by the non-first coefficients and on the special first-block mixing equation, which together enable the ciphertext splice.
3. *Extending the mechanism to EMME.* The unrefreshed-extension argument also applies in Sarkar's EMME framework. The field powers used in EME can be replaced by iterates of a degree-n linear map, because its minimal-polynomial relation provides the corresponding operator identity.

Organization. Section 2 fixes notation and security conventions. Section 3 recalls the EME core in sufficient detail. Section 4 gives the $m = n + 1$ attack and correctness analysis. Section 5 compares it with the original $m = n + 2$ attack. Section 6 extends the attack to general linear maps. Section 7 explains why EME* and EME2-AES avoid the unrefreshed boundary. Section 8 concludes.

2 Preliminaries

Let $n \geq 1$ be the block length in bits. We identify n-bit strings with the finite field $\mathbb{F}_{2^n}$. Addition in $\mathbb{F}_{2^n}$ is written $\oplus$. Let $\alpha \in \mathbb{F}_{2^n}$ denote the element corresponding

to x in a polynomial-basis representation. In EME, multiplication by α is used to generate successive masks.

Let $[r..s]$ denote the integer interval $\{r, r+1, \ldots, s\}$, and write vectors in $\mathbb{F}_{2^n}^m$ as $P = (P_1, \ldots, P_m)$. The integer m is the number of n-bit blocks in the data string; throughout, all block indices start at 1.

Let $E_K : \mathbb{F}_{2^n} \to \mathbb{F}_{2^n}$ be the keyed block-cipher permutation after fixing the underlying block-cipher key, and write E_K^{-1} for its inverse. We use the same symbol E_K for every block-cipher call in EME; this reflects the original single-key design.

We use the term *tweakable enciphering scheme* (TES) for a keyed, tweakable permutation on each supported message space. For a key space $\mathcal{K}$, tweak space $\mathcal{T}$, and message space $\mathcal{M}$, a TES has encryption and decryption maps

$$\mathcal{E}_K^T : \mathcal{M} \to \mathcal{M}, \qquad \mathcal{D}_K^T = (\mathcal{E}_K^T)^{-1}, \qquad K \in \mathcal{K},\ T \in \mathcal{T}.$$

The key K is secret and is understood to be sampled at random. The tweak T is public and may be chosen by the adversary. Enciphering already means that encryption is a permutation on the relevant message space, so no separate "length-preserving" qualifier is needed. In this paper the relevant message space for the attack is $\mathcal{M} = \mathbb{F}_{2^n}^{n+1}$, and the tweak is represented as one field element because that is the form used by the EME core. When the key is fixed, we suppress K and write simply $\mathcal{E}^T$ and $\mathcal{D}^T$.

In the strong pseudorandom permutation (strong-PRP or SPRP) experiment, the adversary has oracle access to both directions, $\mathcal{E}_K^T$ and $\mathcal{D}_K^T$. In the real world these oracles are implemented by the candidate construction under a random secret key K. In the ideal world, for each supported pair $(T, \mathcal{M})$, the oracle is an independent uniformly random permutation on $\mathcal{M}$ together with its inverse. Formally, a distinguisher $\mathcal{A}$ is a possibly randomized oracle algorithm that outputs a bit $b \in \{0, 1\}$. We say that $\mathcal{A}$ *accepts* when $b = 1$, and write $\Pr_{\text{real}}[\mathcal{A} \Rightarrow 1]$ and $\Pr_{\text{ideal}}[\mathcal{A} \Rightarrow 1]$ for its output-one probabilities in the two worlds. Its distinguishing advantage is

$$\mathrm{Adv}^{\text{sprp}}(\mathcal{A}) = \left| \Pr_{\text{real}}[\mathcal{A} \Rightarrow 1] - \Pr_{\text{ideal}}[\mathcal{A} \Rightarrow 1] \right|.$$

An SPRP distinguisher succeeds when this probability gap is non-negligible. A variable-input-length SPRP (VIL-SPRP) asks for the analogous property simultaneously over all supported message spaces.

A *splice* of two block vectors is a new vector obtained by taking some positions from the first vector and the remaining positions from the second. In the attack, a ciphertext splice means exactly such a blockwise recombination of two valid ciphertexts before making a decryption query. Throughout the paper, *unrefreshed extension* means applying the original EME mask schedule beyond its proved range without starting a new chunk or recomputing a fresh global mask. Thus phrases such as *direct, unrefreshed EME-core extension*, *unrefreshed EME core*, *old-mask extension*, and the *old-mask* continuation all refer to the same no-refresh object.

The attack uses one fixed tweak T and the single message space $\mathcal{M} = \mathbb{F}_{2^n}^{n+1}$. It can therefore be read as an ordinary strong-PRP distinguisher on that domain.

For probability estimates, write

$$N = |\mathbb{F}_{2^n}^{n+1}| = 2^{n(n+1)}.$$

3 The EME Core

We recall the EME encryption algorithm in the notation used throughout the paper. The original EME specification supports m-block messages only for $m \in [1..n]$ [7]. We also write the same formula for $m = n + 1$, solely to define the *unrefreshed extension* studied in Section 4.

For a plaintext vector

$$P = (P_1, \ldots, P_m) \in \mathbb{F}_{2^n}^m$$

and tweak $T \in \mathbb{F}_{2^n}$, define $L = \alpha E_K(0^n)$ and

$$a_i = \alpha^{i-1} \qquad (1 \le i \le m).$$

The EME encryption proceeds as follows.

$$\mathsf{PP}_i = P_i \oplus a_i L, \tag{1}$$
$$\mathsf{PPP}_i = E_K(\mathsf{PP}_i), \tag{2}$$
$$\mathsf{MP} = T \oplus \bigoplus_{i=1}^{m} \mathsf{PPP}_i, \tag{3}$$
$$\mathsf{MC} = E_K(\mathsf{MP}), \tag{4}$$
$$M = \mathsf{MP} \oplus \mathsf{MC}, \tag{5}$$
$$\mathsf{CCC}_i = \mathsf{PPP}_i \oplus a_i M, \qquad 2 \le i \le m, \tag{6}$$
$$\mathsf{CCC}_1 = \mathsf{MC} \oplus T \oplus \bigoplus_{i=2}^{m} \mathsf{CCC}_i, \tag{7}$$
$$\mathsf{CC}_i = E_K(\mathsf{CCC}_i), \tag{8}$$
$$C_i = \mathsf{CC}_i \oplus a_i L. \tag{9}$$

The inverse algorithm reverses these steps. In particular, after undoing the second mask and the second ECB pass, it computes

$$\mathsf{MC} = T \oplus \bigoplus_{i=1}^{m} \mathsf{CCC}_i, \qquad \mathsf{MP} = E_K^{-1}(\mathsf{MC}), \qquad M = \mathsf{MP} \oplus \mathsf{MC},$$

then recovers $\mathsf{PPP}_i = \mathsf{CCC}_i \oplus a_i M$ for $i \ge 2$ and

$$\mathsf{PPP}_1 = T \oplus \mathsf{MP} \oplus \bigoplus_{i=2}^{m} \mathsf{PPP}_i.$$

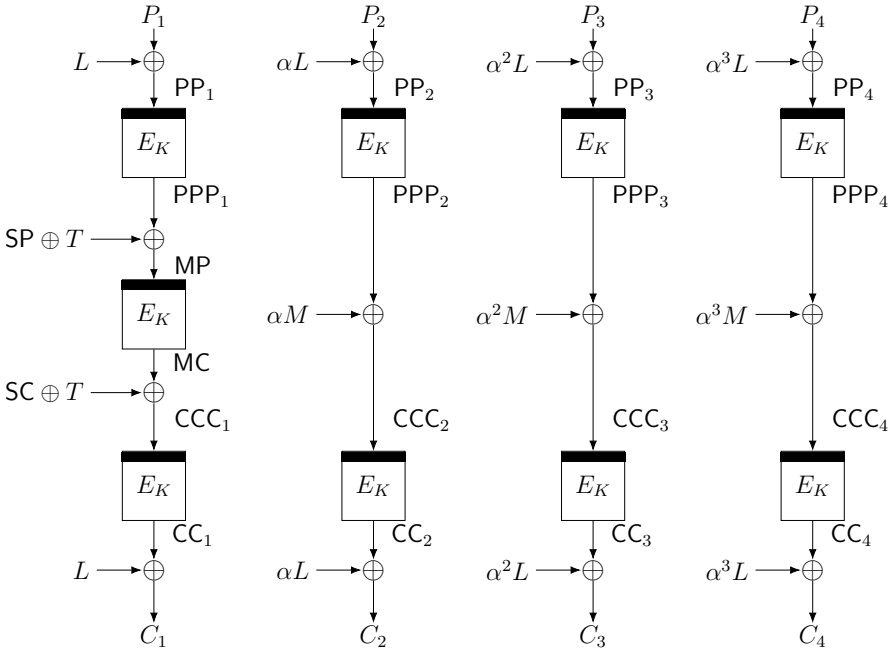


Fig. 1. The original EME structure and its local variables. The whitening mask is $L = \alpha E_K(0^n)$, and the displayed powers $L, \alpha L, \alpha^2 L, \alpha^3 L$ stand for the position masks $a_i L$. After the first ECB pass, $\mathsf{SP} = \mathsf{PPP}_2 \oplus \mathsf{PPP}_3 \oplus \mathsf{PPP}_4$, so the first-column mixer computes $\mathsf{MP} = \mathsf{PPP}_1 \oplus \mathsf{SP} \oplus T$, $\mathsf{MC} = E_K(\mathsf{MP})$, and $M = \mathsf{MP} \oplus \mathsf{MC}$. For the non-first columns, $\mathsf{CCC}_i = \mathsf{PPP}_i \oplus a_i M$; in the displayed four-block instance this means masks $\alpha M, \alpha^2 M, \alpha^3 M$. Finally, $\mathsf{SC} = \mathsf{CCC}_2 \oplus \mathsf{CCC}_3 \oplus \mathsf{CCC}_4$, and the special first post-mixing block is $\mathsf{CCC}_1 = \mathsf{MC} \oplus \mathsf{SC} \oplus T$.

Why the coefficient condition matters. For $m \le n$, the coefficients $a_1, \ldots, a_m$ are linearly independent over $\mathbb{F}_2$. In the proof of EME, accidental collision events reduce to nontrivial affine equations whose leading coefficient is a nonempty sum $\bigoplus_{i \in I} a_i$, which is nonzero for $I \subseteq [1..m]$. Sarkar's generalization makes this proof dependency explicit, replacing the powers of α by iterates of a linear map ψ and proving the corresponding nonzero-sum lemma for $I \subseteq [1..n]$ [12].

4 A Distinguisher at $m = n + 1$

We now define and analyze the attack against the unrefreshed EME-core extension at $m = n+1$. The object under attack is obtained by applying Eqs. (1)–(9) to $n+1$ blocks with the same global mask mechanism. All queries use a single tweak, so by the convention of Section 2 the attack is a standard strong-PRP distinguishing experiment on the fixed domain $\mathbb{F}_{2^n}^{n+1}$.

4.1 The Boundary Relation

At $m = n + 1$, the non-first coefficients

$$a_2, a_3, \ldots, a_{n+1} = \alpha, \alpha^2, \ldots, \alpha^n$$

form a basis of $\mathbb{F}_{2^n}$ over $\mathbb{F}_2$: multiplication by the nonzero element α maps the usual polynomial basis $1, \alpha, \ldots, \alpha^{n-1}$ bijectively to these n elements. Hence there is a unique nonempty subset $S \subseteq \{2, \ldots, n+1\}$ such that

$$\bigoplus_{k \in S} a_k = a_1 = 1. \tag{10}$$

Let

$$J = [1..n+1] \setminus S.$$

Then $1 \in J$.

Why not $m = n$. At $m = n$, the non-first coefficients are only

$$\alpha, \alpha^2, \ldots, \alpha^{n-1},$$

which span an at-most-$(n-1)$-dimensional subspace. In the usual polynomial basis setting, this subspace does not contain 1, and in any case it does not provide the non-first-block basis used in Eq. (10). For $m = n$, the original nonzero-sum condition still applies, since this length remains inside the specified EME domain. Thus $m = n + 1$ is the first point at which the non-first coefficients themselves form an n-element basis and can support the compensation relation $\bigoplus_{k \in S} a_k = 1$.

The adversary is allowed both encryption and decryption queries under the same fixed tweak, as in the strong-PRP experiment.

Theorem 1 (Boundary distinguisher for the unrefreshed EME-core extension). *Consider the direct, unrefreshed EME-core extension defined in Section 2, obtained by using Eqs. (1)–(9) on $m = n+1$ blocks under one fixed tweak T. Let $S \subseteq \{2, \ldots, n+1\}$ be the set satisfying Eq. (10), and let $J = [1..n+1] \setminus S$. There is a strong-PRP distinguisher $\mathcal{A}$ making two encryption queries and one decryption query such that*

$$\Pr_{\text{real}}[\mathcal{A} \Rightarrow 1] = 1$$

and, against an ideal random permutation on $\mathbb{F}_{2^n}^{n+1}$,

$$\Pr_{\text{ideal}}[\mathcal{A} \Rightarrow 1] \leq \frac{2^{n|S|} - 1}{2^{n(n+1)} - 1} + \frac{2^{n(n+1-|S|)}}{2^{n(n+1)} - 2}.$$

In particular, for fixed nonempty S, the ideal-world output-one probability is $O(2^{-n \min\{|S|,|J|\}})$. The statement concerns only the direct, unrefreshed EME-core extension beyond the specified $m \leq n$ EME domain.

Remark 1 (AES polynomial example). For the polynomial $x^{128}+x^7+x^2+x+1$, one has

$$\alpha^{128} \oplus \alpha^7 \oplus \alpha^2 \oplus \alpha = 1.$$

For $n = 128$, this gives

$$S = \{129, 8, 3, 2\}.$$

The relation is one block shorter than the original $m = n + 2$ zero-sum example

$$\{130, 9, 4, 3, 2\},$$

which is obtained by multiplying the above relation by α.

4.2 The Attack

The adversary $\mathcal{A}$ uses a fixed tweak T. The chosen plaintexts should differ only on the positions in S, and in a concrete experiment may be obtained by selecting any nonzero difference vector supported on S.

Step 1. Choose two distinct plaintext vectors $P^{(0)}, P^{(1)} \in \mathbb{F}_{2^n}^{n+1}$ such that

$$P_j^{(0)} = P_j^{(1)} \quad \text{for all } j \in J,$$

and the two vectors differ in at least one position in S.

Step 2. Query encryption twice:

$$C^{(0)} = \mathcal{E}^T(P^{(0)}), \qquad C^{(1)} = \mathcal{E}^T(P^{(1)}).$$

Step 3. Form the spliced ciphertext $C^\star$ by

$$C_j^\star = \begin{cases} C_j^{(0)}, & j \in J, \\ C_j^{(1)}, & j \in S. \end{cases}$$

Step 4. Query decryption:

$$P^\star = \mathcal{D}^T(C^\star).$$

Step 5. Output 1, equivalently accept, if

$$P_k^\star = P_k^{(1)} \qquad \text{for every } k \in S,$$

and output 0 otherwise.

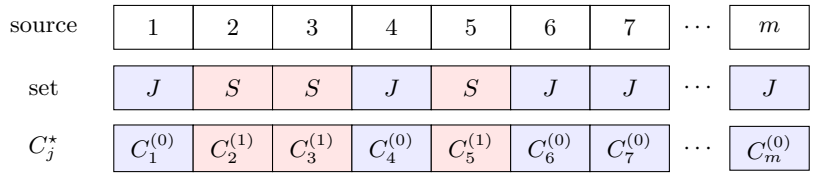


Fig. 2. The splicing pattern used by the distinguisher. The attack is not a zero-sum attack inside the non-first positions; it is a first-block compensation attack using $\bigoplus_{k\in S} a_k = 1$.

4.3 Correctness

We now prove that the spliced ciphertext is accepted in the real EME computation. Decrypting the spliced ciphertext $C^\star$ forces the same internal values MC, MP, and M as in the second encryption query. Once this happens, every copied block indexed by S decrypts to the corresponding block of $P^{(1)}$, so $\mathcal{A}$ outputs 1.

Let

$$\mathsf{PPP}_i^{(b)}, \quad \mathsf{CCC}_i^{(b)}, \quad \mathsf{MP}^{(b)}, \mathsf{MC}^{(b)}, M^{(b)}$$

denote the internal variables in the encryption of $P^{(b)}$, for $b \in \{0, 1\}$. Let

$$\Delta Z = Z^{(1)} \oplus Z^{(0)}$$

for any such variable Z. Since $P_j^{(0)} = P_j^{(1)}$ for all $j \in J$, we have

$$\Delta\mathsf{PPP}_j = 0 \qquad (j \in J).$$

Thus

$$\Delta\mathsf{MP} = \bigoplus_{i=1}^{m} \Delta\mathsf{PPP}_i = \bigoplus_{k\in S} \Delta\mathsf{PPP}_k. \tag{11}$$

Moreover,

$$\Delta M = \Delta\mathsf{MP} \oplus \Delta\mathsf{MC}.$$

For $i \geq 2$, Eq. (6) gives

$$\Delta\mathsf{CCC}_i = \Delta\mathsf{PPP}_i \oplus a_i \Delta M.$$

In particular, for $j \in J \setminus \{1\}$,

$$\Delta\mathsf{CCC}_j = a_j \Delta M. \tag{12}$$

Now consider the first block. By Eq. (7),

$$\Delta\mathsf{CCC}_1 = \Delta\mathsf{MC} \oplus \bigoplus_{i=2}^{m} \Delta\mathsf{CCC}_i.$$

Using $\Delta\mathsf{CCC}_i = \Delta\mathsf{PPP}_i \oplus a_i \Delta M$ for $i \geq 2$, and using Eq. (11), we get

$$\begin{aligned}\Delta\mathsf{CCC}_1 &= \Delta\mathsf{MC} \oplus \bigoplus_{i=2}^{m} \Delta\mathsf{PPP}_i \oplus \left(\bigoplus_{i=2}^{m} a_i\right) \Delta M \\ &= \Delta\mathsf{MC} \oplus \Delta\mathsf{MP} \oplus \left(\bigoplus_{i=2}^{m} a_i\right) \Delta M \\ &= \left(a_1 \oplus \bigoplus_{i=2}^{m} a_i\right) \Delta M,\end{aligned}$$

where $a_1 = 1$.

Let $\mathsf{CCC}^\star$ be the value obtained after undoing the second ECB pass and the output masks of $C^\star$. Since $C_j^\star = C_j^{(0)}$ for $j \in J$ and $C_j^\star = C_j^{(1)}$ for $j \in S$, we have

$$\mathsf{CCC}_j^{(1)} \oplus \mathsf{CCC}_j^\star = \begin{cases} \Delta\mathsf{CCC}_j, & j \in J, \\ 0, & j \in S. \end{cases}$$

Decryption computes $\mathsf{MC}^\star = T \oplus \bigoplus_{i=1}^{m} \mathsf{CCC}_i^\star$. Therefore

$$\begin{aligned}\mathsf{MC}^{(1)} \oplus \mathsf{MC}^\star &= \bigoplus_{j \in J} \Delta\mathsf{CCC}_j \\ &= \Delta\mathsf{CCC}_1 \oplus \bigoplus_{j \in J \setminus \{1\}} a_j \Delta M \\ &= \left(a_1 \oplus \bigoplus_{i=2}^{m} a_i \oplus \bigoplus_{j \in J \setminus \{1\}} a_j\right) \Delta M \\ &= \left(a_1 \oplus \bigoplus_{k \in S} a_k\right) \Delta M \\ &= 0\end{aligned}$$

by Eq. (10). Hence $\mathsf{MC}^\star = \mathsf{MC}^{(1)}$. Since $\mathsf{MC} = E_K(\mathsf{MP})$ and E_K is a permutation, decryption gives

$$\mathsf{MP}^\star = \mathsf{MP}^{(1)} \quad \text{and} \quad M^\star = M^{(1)}.$$

For each $k \in S$, the spliced ciphertext uses the k-th block from $C^{(1)}$, so $\mathsf{CCC}_k^\star = \mathsf{CCC}_k^{(1)}$. Hence

$$\mathsf{PPP}_k^\star = \mathsf{CCC}_k^\star \oplus a_k M^\star = \mathsf{CCC}_k^{(1)} \oplus a_k M^{(1)} = \mathsf{PPP}_k^{(1)}.$$

The first ECB inverse and the input mask are deterministic and identical in the two computations, so $P_k^\star = P_k^{(1)}$ for every $k \in S$. Thus $\mathcal{A}$ always outputs 1 in the real world, proving the real-world part of Theorem 1.

In the ideal world, the same bit-output algorithm outputs 1 only with negligible probability. First consider the event $C^\star = C^{(1)}$. Conditional on $C^{(0)}$, the second encryption output $C^{(1)}$ is uniform over the remaining $N-1$ vectors. The equality $C^\star = C^{(1)}$ holds exactly when $C_j^{(0)} = C_j^{(1)}$ for every $j \in J$. Since $C^{(1)} \neq C^{(0)}$, at most $2^{n|S|} - 1$ choices of $C^{(1)}$ have this property. Thus

$$\Pr[C^\star = C^{(1)}] \leq \frac{2^{n|S|} - 1}{N - 1}.$$

The event $C^\star = C^{(0)}$ cannot make $\mathcal{A}$ output 1: the resulting plaintext would be $P^{(0)}$, and $P^{(0)}$ differs from $P^{(1)}$ in at least one position in S.

Condition now on $C^\star \notin \{C^{(0)}, C^{(1)}\}$. Then $\mathcal{D}^T(C^\star)$ is uniform over the $N-2$ messages not already used as encryption queries. At most $2^{n(n+1-|S|)}$ messages satisfy the $|S|$ prescribed block equalities $P_k^\star = P_k^{(1)}$ for every $k \in S$. Hence

$$\Pr_{\text{ideal}}[\mathcal{A} \Rightarrow 1] \leq \frac{2^{n|S|} - 1}{2^{n(n+1)} - 1} + \frac{2^{n(n+1-|S|)}}{2^{n(n+1)} - 2}.$$

Because $|S| + |J| = n+1$, the first term is $O(2^{-n|J|})$ and the second is $O(2^{-n|S|})$, giving the simplified bound $O(2^{-n \min\{|S|,|J|\}})$. This proves Theorem 1.

5 Comparison with the Original $m = n + 2$ Attack

The boundary attack differs from the original length-restriction attack in [7]. The original attack assumes a nonempty proper set $J_0 \subseteq \{2, \ldots, m\}$ such that

$$\bigoplus_{j \in J_0} a_j = 0.$$

It chooses two plaintexts that differ only in the first block, splices ciphertext blocks according to J_0, and thereby forces a selected MC value to recur. For AES with the standard polynomial $x^{128} + x^7 + x^2 + x + 1$, the displayed relation is

$$\alpha^{129} \oplus \alpha^8 \oplus \alpha^3 \oplus \alpha^2 \oplus \alpha = 0,$$

which requires index 130, namely $m \geq 130 = n + 2$.

At $m = n + 1$, no such nonempty zero-sum subset exists among $\alpha, \ldots, \alpha^n$, because these n elements form a basis. The boundary attack uses a different relation:

$$\bigoplus_{k \in S} a_k = 1.$$

It then uses $1 \in J$ and the special definition of CCC_1 to contribute the missing $a_1 = 1$ term. The original attack is therefore a *non-first-block zero-sum attack*, whereas the boundary attack is a *first-block compensation attack*. The first dependence does not appear as a zero-sum among non-first coefficients; it appears as a representation of the missing first coefficient.

Feature	Original $m \geq n+2$ attack	Present $m = n+1$ attack
Algebraic relation	$\bigoplus_{j \in J_0} a_j = 0$	$\bigoplus_{k \in S} a_k = 1$
Index set	$J_0 \subseteq \{2, \ldots, m\}$	$S \subseteq \{2, \ldots, m\}$, with $1 \in J$
Plaintexts	differ only in block 1	agree on J, differ inside S
Use of CCC_1	not essential	essential
Earliest length	$n+2$ blocks	$n+1$ blocks

Table 1. The boundary attack is not the original attack with a smaller index. It uses the first block's special equation as a compensation term.

6 Generalization to Linear-Map EMME

Sarkar proposed a general mixing strategy, called EMME here, replacing multiplication by α with a linear map

$$\psi : \mathbb{F}_{2^n} \to \mathbb{F}_{2^n}$$

whose minimal polynomial over $\mathbb{F}_2$ is primitive of degree n [12]. The EME position masks are generated by applying the same operator to the current mask value: each scalar product $a_i X = \alpha^{i-1} X$, for $X \in \{L, M\}$, is replaced by

$$\psi^{i-1}(X).$$

In this section, $\oplus$ and $\bigoplus$ between linear operators denote pointwise addition of maps over $\mathbb{F}_2$.

Sarkar explicitly observes that the EME proof uses the nonzero-sum property of the powering method and proves the corresponding lemma

$$\bigoplus_{i \in I} \psi^{i-1}(M) \neq 0 \quad \text{for nonempty } I \subseteq \{1, \ldots, n\},\ M \neq 0.$$

This lemma gives the proof for the valid range $m \leq n$. The next dependence, forced exactly at degree n, yields the same boundary attack for the direct unrefreshed EMME-core extension.

We separate the condition used for the proof from the condition needed for the attack. Sarkar's sufficient condition for inheriting the EME proof is the primitive degree-n minimal polynomial assumption, which gives the needed nonzero-sum property in the first n positions. The attack uses only the weaker boundary identity expressing Id as a nonempty sum of non-first iterates $\psi, \ldots, \psi^n$. A degree-n minimal polynomial with nonzero constant term supplies such an identity.

Let the minimal polynomial of ψ be

$$\tau(x) = x^n \oplus c_{n-1} x^{n-1} \oplus \cdots \oplus c_1 x \oplus c_0, \tag{13}$$

where $c_0 = 1$ in the cases considered here: primitive polynomials have nonzero constant term, and Theorem 2 assumes the same condition over $\mathbb{F}_2$. When τ is

evaluated at ψ, the constant term $c_0 = 1$ becomes the identity map Id. Therefore $\tau(\psi) = 0$ gives

$$\psi^n \oplus \bigoplus_{\substack{1 \le r \le n-1 \\ c_r = 1}} \psi^r \oplus \mathsf{Id} = 0.$$

Equivalently,

$$\mathsf{Id} = \psi^n \oplus \bigoplus_{\substack{1 \le r \le n-1 \\ c_r = 1}} \psi^r. \tag{14}$$

Define

$$S_\psi = \{n+1\} \cup \{r+1 : 1 \le r \le n-1,\ c_r = 1\} \subseteq \{2, \dots, n+1\}.$$

Then Eq. (14) says

$$\mathsf{Id} = \bigoplus_{k \in S_\psi} \psi^{k-1}.$$

Here we use the same unrefreshed-extension convention as in Section 2.

Theorem 2 (Linear-map generalization). *Consider the direct, unrefreshed EMME-core extension obtained from Eqs. (1)–(9) on $m = n+1$ blocks by replacing each scalar mask $a_i X$, for $X \in \{L, M\}$, with $\psi^{i-1}(X)$, under one fixed tweak T. Suppose that ψ has a degree-n minimal polynomial with nonzero constant term, and let $S_\psi \subseteq \{2, \dots, n+1\}$ be defined as above. There is a strong-PRP distinguisher $\mathcal{A}$ making two encryption queries and one decryption query such that*

$$\Pr_{\text{real}}[\mathcal{A} \Rightarrow 1] = 1$$

and, against an ideal random permutation on $\mathbb{F}_{2^n}^{n+1}$,

$$\Pr_{\text{ideal}}[\mathcal{A} \Rightarrow 1] \le \frac{2^{n|S_\psi|} - 1}{2^{n(n+1)} - 1} + \frac{2^{n(n+1-|S_\psi|)}}{2^{n(n+1)} - 2}.$$

In particular, with $J_\psi = [1..n+1] \setminus S_\psi$, the ideal-world output-one probability is

$$O(2^{-n \min\{|S_\psi|, |J_\psi|\}}).$$

The statement concerns only this direct, unrefreshed EMME-core extension beyond the proof domain; refreshed arbitrary-length designs are outside its scope.

Proof. The proof is identical to the correctness analysis in Section 4.3, with scalars a_i replaced by linear operators $A_i = \psi^{i-1}$ and with $a_1 = 1$ replaced by $A_1 = \mathsf{Id}$. For the two encryption queries, choose plaintexts equal on $J_\psi = [1..n+1] \setminus S_\psi$. Then, for $j \in J_\psi \setminus \{1\}$,

$$\Delta\mathsf{CCC}_j = \psi^{j-1}(\Delta M).$$

The first block contributes

$$\Delta\mathsf{CCC}_1 = \left(\mathsf{Id} \oplus \bigoplus_{i=2}^{n+1} \psi^{i-1}\right)(\Delta M).$$

The spliced ciphertext satisfies

$$\begin{aligned}\mathsf{MC}^{(1)} \oplus \mathsf{MC}^{\star} &= \left(\mathsf{Id} \oplus \bigoplus_{k \in S_\psi} \psi^{k-1}\right)(\Delta M) \\ &= 0\end{aligned}$$

by Eq. (14). Therefore $\mathsf{MC}^{\star} = \mathsf{MC}^{(1)}$, $\mathsf{MP}^{\star} = \mathsf{MP}^{(1)}$, and $M^{\star} = M^{(1)}$. The blocks indexed by S_ψ are copied from $C^{(1)}$, so the recovered plaintext blocks at these positions equal those of $P^{(1)}$. Hence $\mathcal{A}$ outputs 1 in the real world with probability one. The ideal-world analysis is identical to the proof of Theorem 1, with S replaced by S_ψ.

Remark 2. The statement is consistent with Sarkar's security claim: the EMME proof, like the EME proof it generalizes, uses the nonzero-sum lemma only for index sets within the first n powers. The attack uses the first linear dependence forced by the degree-n minimal polynomial and therefore sits exactly outside that proof domain.

7 Relation to EME* and IEEE EME2-AES

The attack is a boundary result for the *unrefreshed extension* defined in Section 2. That object is distinct from EME* and EME2-AES.

EME* was introduced by Halevi to remove the original limitations of EME: the original EME applies only to integral numbers of blocks and at most n blocks, whereas EME* supports almost arbitrary bit lengths and associated data [8]. Earlier EME+ already suggested dividing long inputs into chunks of at most n blocks and using a refreshed mask for later chunks [7]. In particular, at the first block after a chunk boundary, the construction computes a fresh MP, MC, and M for the new chunk, instead of continuing with the old-mask term $a_{n+1}M$, i.e., the no-refresh continuation of the previous global mask.

IEEE 1619.2-2025 specifies EME2-AES as an AES-based wide-block encryption with associated data for shared storage media [9]. For standardization context, public IEEE catalogue descriptions of earlier IEEE 1619.2 revisions listed both EME2-AES and XCB-AES, while the public entry for IEEE 1619.2-2025 describes the EME2-AES mode [9]. In EME2-AES, the mask refresh across chunks removes the old-mask continuation used by our splice.

The standard is not just a presentation of the original EME core: it fixes AES as the narrow block cipher, defines byte-level associated data processing, and

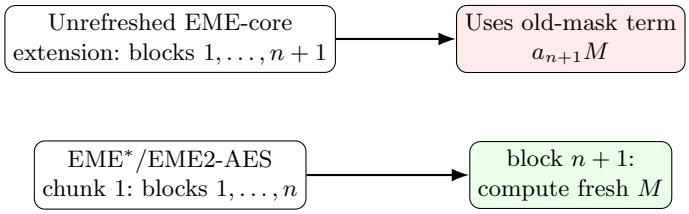


Fig. 3. The attack applies to the unrefreshed EME-core extension of the first EME chunk. EME* and EME2-AES cross the boundary by computing a fresh mask, so the algebraic relation used in the attack no longer matches the actual mixing equations.

specifies an arbitrary-byte-length data-unit interface for storage. In particular, the encryption procedure refreshes the mask when the loop reaches the boundary

$$(j-1) \bmod 128 = 0,$$

rather than continuing to use the current mask, as specified by the standard's pseudocode and illustrated for 130 full blocks and one partial block [9].

The $m = n+1$ distinguisher is therefore a design-boundary result:

do not extend the original EME core past n blocks without refreshing.

It does not break EME* or EME2-AES as specified. It gives an algebraic explanation for the role of their chunking and mask-refresh mechanisms and for why the standardized EME2-AES design avoids the unrefreshed boundary considered here.

8 Conclusion

The original EME proof and specification stop at $m \leq n$. Halevi and Rogaway showed that the same direct, unrefreshed EME-core extension to $m \geq n+2$ admits a simple distinguisher, but left the exact $m = n+1$ boundary unresolved. This extension is already distinguishable at $m = n+1$. The attack exploits not a zero-sum among the non-first coefficients, but the fact that those n coefficients form a basis and therefore represent the first coefficient. The first block's special equation then turns this representation into a full cancellation of $\mathsf{MC}^{(1)} \oplus \mathsf{MC}^{\star}$.

The attack generalizes to EMME and analogous linear-map replacements whose minimal polynomial has degree n: the minimal polynomial supplies the operator identity needed to represent Id by non-first iterates. The phenomenon is not tied to one polynomial basis or one field representation.

For implementations and standards, the result clarifies the role of mask refresh rather than challenging the refreshed designs themselves. These arbitrary-length variants avoid the boundary by refreshing the mask after at most n blocks.

The result should be viewed as an algebraic explanation of the refresh rule: carrying the first EME mask unchanged into block $n+1$ creates a splicing relation.

References

1. A. S. Bhati and E. Andreeva. Breaking the IEEE encryption standard: XCB-AES in two queries. In Y. T. Kalai and S. F. Kamara, editors, *Advances in Cryptology – CRYPTO 2025*, Part V, Lecture Notes in Computer Science, vol. 16004, pp. 172–199. Springer, Cham, 2025. https://doi.org/10.1007/978-3-032-01901-1_6.
2. Y. L. Chen, M. Davidson, M. Dworkin, J. Kelsey, Y. Sasaki, M. Sönmez Turan, D. Chang, N. Mouha, and A. Thompson. *Requirements for Cryptographic Accordions*. NIST Interagency/Internal Report (NIST IR) 8552, National Institute of Standards and Technology, Gaithersburg, MD, April 2025. https://doi.org/10.6028/NIST.IR.8552.
3. P. Crowley, N. Huckleberry, and E. Biggers. Length-preserving encryption with HCTR2. Cryptology ePrint Archive, Paper 2021/1441, 2021; revised 2023. https://eprint.iacr.org/2021/1441.
4. N. Datta, A. Dutta, S. Ghosh, E. List, and H. Nandi. HCTR+: an optimally secure TBC-based accordion mode. *IACR Transactions on Symmetric Cryptology*, 2025(3):183–229, 2025. https://doi.org/10.46586/tosc.v2025.i3.183-229.
5. C. Dobraunig, K. Matusiewicz, B. Mennink, and A. Tereschenko. Efficient instances of docked double decker with AES, and application to authenticated encryption. In *Advances in Cryptology – EUROCRYPT 2025*, Part I, pp. 62–92. Springer, Cham, 2025. https://doi.org/10.1007/978-3-031-91107-1_3.
6. S. Halevi and P. Rogaway. A tweakable enciphering mode. In D. Boneh, editor, *Advances in Cryptology – CRYPTO 2003*, Lecture Notes in Computer Science, vol. 2729, pp. 482–499. Springer, Berlin, Heidelberg, 2003. https://doi.org/10.1007/978-3-540-45146-4_28.
7. S. Halevi and P. Rogaway. A parallelizable enciphering mode. In T. Okamoto, editor, *Topics in Cryptology – CT-RSA 2004*, Lecture Notes in Computer Science, vol. 2964, pp. 292–304. Springer, Berlin, Heidelberg, 2004. https://doi.org/10.1007/978-3-540-24660-2_23.
8. S. Halevi. EME*: extending EME to handle arbitrary-length messages with associated data. In A. Canteaut and K. Viswanathan, editors, *Progress in Cryptology – INDOCRYPT 2004*, Lecture Notes in Computer Science, vol. 3348, pp. 315–327. Springer, Berlin, Heidelberg, 2004. https://doi.org/10.1007/978-3-540-30556-9_25.
9. IEEE Computer Society. *IEEE Std 1619.2-2025: IEEE Standard for Wide-Block Encryption for Shared Storage Media*. IEEE, approved 10 September 2025; published 20 November 2025. https://standards.ieee.org/ieee/1619.2/11949/.
10. B. Lee. A BBB-secure accordion mode from HCTR. Presentation at the NIST Workshop on the Requirements for an Accordion Cipher Mode, Gaithersburg, MD, June 2024. https://csrc.nist.gov/Presentations/2024/a-bbb-secure-accordion-mode-from-hctr.
11. D. A. McGrew and S. R. Fluhrer. The security of the Extended Codebook (XCB) mode of operation. In C. Adams, A. Miri, and M. Wiener, editors, *Selected Areas in Cryptography – SAC 2007*, Lecture Notes in Computer Science, vol. 4876, pp. 311–327. Springer, Berlin, Heidelberg, 2007. https://doi.org/10.1007/978-3-540-77360-3_20.

12. P. Sarkar. A general mixing strategy for the ECB-Mix-ECB mode of operation. *Information Processing Letters*, 109(2):121–123, 2008. https://doi.org/10.1016/j.ipl.2008.09.012.
13. A. Thompson and M. Sönmez Turan. *NIST Workshop on the Requirements for an Accordion Cipher Mode 2024: Workshop Report*. NIST Interagency/Internal Report (NIST IR) 8537, National Institute of Standards and Technology, Gaithersburg, MD, November 2024. https://doi.org/10.6028/NIST.IR.8537.
14. P. Wang, D. Feng, and W. Wu. HCTR: a variable-input-length enciphering mode. In D. Feng, D. Lin, and M. Yung, editors, *Information Security and Cryptology – CISC 2005*, Lecture Notes in Computer Science, vol. 3822, pp. 175–188. Springer, Berlin, Heidelberg, 2005. https://doi.org/10.1007/11599548_15.