# Extended Frege and Gaussian Elimination

Michael Soltys
Department of Computing and Software
McMaster University
1280 Main Street West
Hamilton, Ontario L8S 4K1, CANADA
Email: soltys@mcmaster.ca

2002

**Abstract**

We show that the Gaussian Elimination algorithm can be proven correct with uniform Extended Frege proofs of polynomial size, and hence feasibly. More precisely, we give short uniform Extended Frege proofs of the tautologies that express the following: given a matrix $A$, the Gaussian Elimination algorithm reduces $A$ to row-echelon form. We also show that the consequence of this is that a large class of matrix identities can be proven with short uniform Extended Frege proofs, and hence feasibly.

## 1 Introduction

Gaussian Elimination is a well studied algorithm. It consists in reducing a given matrix to row-echelon form by performing on it a sequence of elementary row operations (such as adding a multiple of one row to another, exchanging two rows, or multiplying a row by a constant).

Since Gaussian Elimination is a polytime algorithm, it was known that it can be proven total in standard logical theories for polynomial time (polytime) reasoning (such as Cook's $\mathbf{PV}$ or Buss' $\mathbf{S}^1_2$). In this paper, we give a direct proof of the correctness of Gaussian Elimination (and hence also of its totality) with Extended Frege (eFrege) proofs of size polynomial in the size of the given matrix. By correctness we mean the following statement: the Gaussian Elimination algorithm reduces a matrix to row-echelon form.

This result is important for several reasons: First of all, it was assumed that Gaussian Elimination is "well behaved"—from a proof complexity point of view; that is, it was assumed that the standard properties of Gaussian Elimination can be proven feasibly. But there are examples of algorithms for which we do not know if they can be proven correct within their complexity class (e.g., Berkowitz's algorithm; see [8]). Thus, while the above mentioned assumption was reasonable, a direct feasible proof of the correctness of Gaussian Elimination was nevertheless desirable.

Second of all, the Gaussian Elimination algorithm is a cornerstone algorithm in linear algebra—the main "engine for computation" in standard textbooks of linear algebra. A substantial portion of matrix algebra can be proven easily (and feasibly) from the correctness of Gaussian Elimination. We show that a class of matrix identities, which we call "hard matrix identities" (because it appears that they do not have polynomial size Frege proofs) follow directly from the proof of correctness of Gaussian Elimination (an example of such an identity is $AB = I \supset BA = I$). Thus, a substantial portion of universal matrix algebra can be proven with short eFrege proofs.

The hard matrix identities bring us to the last point. The separation of Frege and eFrege is a fundamental open problem in theoretical computer science. Cook proposed $AB = I \supset BA = I$ as a candidate for showing this separation, since it appears that this identity does not have polynomial size Frege proofs. On the other hand, it does have polynomial size eFrege proofs (a consequence of our polysize eFrege proofs of correctness of Gaussian Elimination). We hope that an exploration of the proof complexity of matrix algebra might shed some light on the alleged separation of these two proof systems.

We define Frege and eFrege in section 2.1. In section 2.2 we show how to express universal matrix identities with propositional formulas. In section 3 we prove the main result of this paper: we show that the correctness of Gaussian Elimination can be proven with uniform polysize eFrege proofs, and that therefore hard matrix identities also have uniform polysize eFrege proofs.

# 2 Preliminaries

## 2.1 Proof systems, Frege, and eFrege

Proof Complexity is an area of mathematics and theoretical computer science that studies the length of proofs in propositional logic. It is an area of study that is fundamentally connected both to major open questions of computational complexity theory and to practical properties of automated theorem provers ([2]). Let TAUT be the set of all tautologies. A *propositional proof system* is just a polytime predicate $P \subseteq \Sigma^* \times \text{TAUT}$ such that $\phi \in \text{TAUT} \iff \exists x P(x, \phi)$. $P$ is *poly-bounded* if there exists a polynomial $p$ such that:

$$\phi \in \text{TAUT} \iff \exists x(|x| \leq p(|\phi|) \wedge P(x, \phi))$$

The existence of a poly-bounded proof system is related to the fundamental question about complexity classes: P = NP ? Cook and Reckhow proved that NP = co-NP iff there is a poly-bounded proof system for TAUT ([3]). On the other hand, if P = NP then NP = co-NP. Thus, if there is no poly-bounded proof system, then NP $\neq$ co-NP, and that implies that P $\neq$ NP.

Unfortunately, a proof system is such a general object (just a polytime predicate, as defined above), that it is hopeless at the moment to show directly that there is no polybounded proof system (if that is indeed the case, because it is possible—but not widely believed—that NP = co-NP, but P $\neq$ NP). Instead, the program proposed by Cook is to show lower bounds for the common propositional proofs systems (such as resolution, Frege systems, etc.), of increasing strength. This is a good approach, because lower bounds for propositional proof systems are of interest independently of the P = NP ? question. In particular, they are of interest to automated reasoning in artificial intelligence, and to lower bounds for algorithms for satisfiability (see [2] for more details).

In Figure 1 we show a table of the principal propositional proof systems. Exponential lower bounds exist for the proof systems below the line. The strongest propositional proof system (Quantified Frege) is shown in the top, and the weakest (Truth Tables) is shown in the bottom. Each system can simulate the one below. The systems Frege and PK are equivalent in the sense that they $p$-simulate each other. We say that a proof system $P$ *p-simulates* a proof system $P'$ if there exists a polytime function $f$ such that $P'(x, \phi)$ holds iff $P(f(x), \phi)$ holds. In other words, all the proofs of $P'$ can be "reproduced" in $P$ with a small increase in size.

| |
|---|
| Quantified Frege |
| Extended Frege, Substitution Frege, Renaming Frege |
| Permutation Frege |
| Frege, PK |
| Bounded Depth (BD) Frege |
| Resolution |
| Truth Tables |

Table 1: Propositional proof systems

As was mentioned above, the program is to show lower bounds for standard proof systems of increasing complexity. So far, lower bounds exist for Resolution (Haken [4] who showed exponential lower bounds for the pigeonhole principle), and Bounded Depth Frege (Ajtai [1] who also showed exponential lower bounds for the pigeonhole principle, but in Bounded Depth Frege—this result formed the basis of much of the research in proof complexity in the following decade, see [2]), but no lower bounds exist for stronger systems. In particular, there is no separation between the Frege and Extended Frege proof systems.

The (alleged) separation between Frege and Extended Frege is a *fundamental open problem* in theoretical computer science. A Frege system is a propositional proof system with finitely many rules (see [9] for details). It was shown in [3], that Frege systems with different rules and over a different basis $p$-simulate each other. Thus, it is a very robust class of proof systems. Extended Frege (eFrege), is Frege with the extension rule. This rule allows the possibility of abbreviating formulas by definitions. Thus, Frege corresponds to reasoning with Boolean formulas, while eFrege corresponds to reasoning with Boolean circuits.

Frege and eFrege are well known systems, and they are studied in depth in [9] and [5]. We assume that Boolean formulas are defined in a standard way with the connectives $\{\vee, \wedge, \neg, \oplus, \supset, \leftrightarrow\}$, and that $0, 1$ are false and true, respectively. A Frege proof is a sequence of Boolean formulas $\{\phi_1, \phi_2, \ldots, \phi_n\}$, where $\phi_n$ is the conclusion (i.e., the tautology which we prove), and each $\phi_i$ is either an axiom of the form $\phi \vee \neg\phi$, or it follows from some $\phi_j$, $j < i$ by a rule. A rule is a $(k+1)$-tuple of formulas written as:

$$\frac{\theta_1, \ldots, \theta_k}{\theta_0}$$

Which rules we choose is immaterial (as long as they are complete and sound) by the already mentioned result of Cook and Reckhow ([3]).

An eFrege proof is a sequence of Boolean formulas $\{\phi_1, \phi_2, \ldots, \phi_n\}$, as before, but now there is a third possibility: a formula $\phi_i$ might be a definition $p \leftrightarrow \theta$, where $p$ is a new atom that does not appear in $\theta$ nor in $\phi_j$, for $j < i$.

A term that might require some clarification is "uniform." We mentioned in the introduction that we show the correctness of Gaussian Elimination with *uniform* polysize eFrege proofs. The correctness of Gaussian Elimination will be stated as a family of tautologies, parametrized by the size of the given matrices (intuitively, the tautologies $\tau_1, \tau_2, \tau_3, \ldots$, express the correctness of Gaussian Elimination for matrices with $1, 2, 3, \ldots$, rows, respectively). Each tautology $\tau_n$ has an eFrege proof of size bounded by a fixed polynomial in $n$, and each proof can be generated *uniformly* (in polytime); that is, the proofs are not wildly different, but have a similar structure. The uniformity condition is important, since uniform polysize eFrege proofs provide *feasible* proofs, while polysize eFrege proofs alone do not necessarily provide feasible proofs. We sometimes abuse notation, and abbreviate "uniform polysize" by "short." Finally, the uniformity of the derivations will be obvious, so, as a rule, we will not point it out.

## 2.2 Expressing matrix identities

In [6], the author designed a quantifier-free, three sorted (where the sorts are indices, field elements, and matrices) logical theory for linear algebra, called LA. In LA it is possible to express universal matrix identities, such as for example $AB = I \supset BA = I$, and also prove all the ring properties of matrices (associativity of matrix addition and multiplication, commutativity of addition, etc.). Then, it was also shown how to translate a formula in the language of LA, to a family of propositional formulas, where the parameters of the translation were the sizes of matrices in the formula. Hence, since the general problem of expressing matrix identities as tautologies has been solved in [6], here we just give an outline that is enough for our purposes. (As an aside, note that LA *cannot* formalize Gaussian Elimination, so we cannot just take it and use it here; we really need extension definitions in order to formalize Gaussian Elimination.)

Matrices have entries from some field. We assume that the underling field is $\mathbb{Z}_2 = \{0, 1\}$, the field of two elements. An $n \times m$ matrix $A$ over the field $\mathbb{Z}_2$ can be easily represented with $nm$ Boolean variables $A_{11}, A_{12}, \ldots, A_{nm}$. For a bigger field, we need to encode each entry of $A$ by several Boolean variables, and the

Boolean simulation of field operations is technically more involved. However, all the results in this paper hold for bigger fields as well, so without loss of generality, we can restrict ourselves to the field $\mathbb{Z}_2$.

If $a, b$ are field variables over $\mathbb{Z}_2$, then $a \cdot b$ can be represented by the Boolean formula $a \wedge b$, and $a + b$ can be represented by $a \oplus b$. For a bigger field, and a thorough study of the relation between algebraic expressions and Boolean formulas, see [10].

As was mentioned above, we associate an $n \times m$ matrix $A$ over $\mathbb{Z}_2$ with $nm$ Boolean variables $A_{ij}$. To express the usual matrix terms ($A + B$, $A(B + D)$, etc.), we use extension definitions. For example, to express $A + B$ we introduce a new set of Boolean variables, $C_{ij}$, and define them as follows:

$$C_{ij} \leftrightarrow (A_{ij} \oplus B_{ij}) \tag{1}$$

for $1 \leq i \leq n, 1 \leq j \leq m$ if $A$ and $B$ are $n \times m$ matrices. In general, $C$ will be used to denote *new* variables. Let $\|C = A + B\|_{n,m}$ denote the *set* of extension definitions given by (1), for all $1 \leq i \leq n$ and $1 \leq j \leq m$. That is, $\|C = A + B\|_{n,m}$ denotes $\{C_{ij} \leftrightarrow (A_{ij} \oplus B_{ij})\}_{1 \leq i \leq n, 1 \leq j \leq m}$.

To express $C = AB$, we define each $C_{ij}$ as:

$$C_{ij} \leftrightarrow ((A_{i1} \wedge B_{1j}) \oplus (A_{21} \wedge B_{2j}) \oplus \ldots \oplus (A_{in} \wedge B_{nj})) \tag{2}$$

Note that our Boolean connectives have fan-in 2, so the right-hand side of the above formula should be parenthesized appropriately; assume that it is parenthesized left to right. In general, assume that whenever we write a formula of the form $\phi_1 \circ \phi_2 \circ \cdots \circ \phi_n$, where "$\circ$" denotes some Boolean connective, we mean its left to right parenthesization, that is, we mean: $\phi_1 \circ (\phi_2 \circ (\cdots \circ \phi_n) \cdots)$

Let $\|C = AB\|_n$ denote the set of extension definitions given by (2), i.e., it denotes $\{C_{ij} \leftrightarrow ((A_{i1} \wedge B_{1j}) \oplus (A_{21} \wedge B_{2j}) \oplus \ldots \oplus (A_{in} \wedge B_{nj}))\}_{1 \leq i,j \leq n}$. Note that the product of two matrices of sizes $n \times p$ and $p \times m$ can be defined by padding the matrices with zeros to make them square of size $\max\{n, p, m\} \times \max\{n, p, m\}$.

We can also define iterated matrix products. Suppose that we want to define the iterated product $A_1 A_2 \cdots A_m$, where all $A_i$ are $n \times n$ matrices. We define $C_1, C_2, \ldots, C_{m-1}$ sequentially as follows: $C_1 = A_1 A_2$, $C_2 = C_1 A_3$, etc., until we obtain $C_{m-1} = C_{m-2} A_m$. Thus, the set of extension definitions that define $C_{m-1}$, the product $A_1 A_2 \cdots A_m$, is the following:

$$\|C_1 = A_1 A_2\|_n, \|C_2 = C_1 A_3\|_n, \ldots, \|C_{m-1} = C_{m-2} A_m\|_n \tag{3}$$

This definition illustrates the interplay between matrix variables, and Boolean variables: each $C_i$ denotes a matrix, in the context of matrix algebra, and a *set* of Boolean variables, in the context of Boolean formulas. While the expression $C_{i+1} = C_i A_{i+1}$ is a matrix identity, $\|C_{i+1} = C_i A_{i+1}\|_n$ is a set of extension definitions that define the set of Boolean variables denoted by $C_{i+1}$, in terms of the sets of Boolean variables that define $C_i$ and $A_{i+1}$. Because this interplay is well defined, we sometimes abuse notation, and go between the two "modes" in the proofs.

If $A, B$ are $n \times m$ matrices, let $\|A = B\|_{n,m}$ denote the following set of extension definitions:

$$\{A_{ij} \leftrightarrow B_{ij}\}_{1 \leq i \leq n, 1 \leq j \leq m} \tag{4}$$

Note that over more general fields, we would also need to define the scalar multiplication of a matrix. But, over $\mathbb{Z}_2$, $aA$ is either the zero matrix, if $a = 0$, or it is $A$ if $a = 1$. In any case, it is easy to define it, even for bigger fields.

Now, we can define more complicated formulas recursively. Suppose that we want to state the following: $A + (B + E) = AE$. First we would express $B + E$ with $C_1$, $A + C_1$ with $C_2$, $AE$ with $C_3$, and finally, we would state $C_2 = C_3$. Here, $C_1, C_2, C_3$ are the *sets of new extension variables.* Thus, $A+(B+E) = AE$ would be expressed as follows:

$$\|C_1 = B + E\|_{n,n}, \|C_2 = A + C_1\|_{n,n}, \|C_3 = AE\|_n, \|C_2 = C_3\|_{n,n}$$

where $n$ is the *parameter* of the translation; there is a Boolean formula for each value of $n$, i.e., for each fixed size of matrices. The matrices are assumed to be square, of size $n$.

As a second example, consider $AB = I \supset BA = I$. First of all, note that $I$ is a *constant* matrix, for any given size $n$. That is, we have a set of extension definitions $\{I_{ij} \leftrightarrow 0\}_{1 \leq i \neq j \leq n}, \{I_{ii} \leftrightarrow 1\}_{1 \leq i \leq n}$. We state the identity as follows:

$$\bigwedge \|AB = I\|_n \supset \bigwedge \|BA = I\|_n \tag{5}$$

Note that $\|AB = I\|_n$ is a set of extension definitions, so $\bigwedge \|AB = I\|_n$ denotes the conjunction (properly parenthesized) of all these extension definitions. Same holds for $\bigwedge \|BA = I\|_n$. We can take the extension definitions for $I$ to be axioms of our eFrege system (instead of adding them to tautology (5), but this does not matter either way for proof length). Again, $n$ is the parameter of the translation.

From this, it is hopefully clear how to translate general (universal) matrix identities into families of Boolean formulas. Thus, instead of writing (5), we will simply state $\|AB = I \supset BA = I\|_n$, and in general, if $\alpha(A_1, A_2, \ldots, A_n)$ is a universal matrix identity, where $A_1, A_2, \ldots, A_n$ denote the free matrix variables, then $\|\alpha(A_1, A_2, \ldots, A_n)\|_m$ will denote the tautology we obtain when all matrices have size $m$, and $\{\|\alpha(A_1, A_2, \ldots, A_n)\|_m\}$ will denote the family of all such tautologies, parametrized by $m$.

The following four matrix identities are allegedly hard for Frege, and so we call them *hard matrix identities* (we explain below exactly what we mean by "hard for Frege").

$$(AB = I \wedge AC = I) \supset B = C \tag{I}$$

$$AB = I \supset (AC \neq 0 \vee C = 0) \tag{II}$$

$$AB = I \supset BA = I \tag{III}$$

$$AB = I \supset A^t B^t = I \tag{IV}$$

Identity I states that right inverses are unique, identity II states that units are not zero-divisors, and identity III states that a right inverse is an inverse.

Identity III was proposed by Cook as a candidate for the separation of Frege and Extended Frege propositional proof systems.

We explain what we mean by "hard for Frege." Consider for example identity III, and let $\{\tau_n\}$ be (5). Then, there is no polynomial $p(x) \in \mathbb{N}[x]$ such that for all $n$, $\tau_n$ has a Frege proof of size bounded by $p(n)$.

**Conjecture 1** *Identities* I, II, III, IV *are hard for Frege; that is, if $\alpha$ is one of* I, II, III, *or* IV, *then, for every polynomial $p \in \mathbb{N}[x]$, there exists an $n_0$ sufficiently big so that $\|\alpha\|_{n_0}$ does* not *have a Frege proof of size $\leq p(n_0)$.*

It is enough to show that one of these identities, e.g., $AB = I \supset BA = I$, cannot be proven in polysize Frege to conclude that none of them can be proven in polysize Frege. If one of them can be proven in polysize Frege (or eFrege), then all can be proven in polysize Frege (or eFrege). See [6] for details.

**Theorem 1** *All the ring properties of matrices can be proven in polysize Frege. That is, commutativity and associativity of matrix addition and multiplication, as well as distributivity, can be proven in polysize Frege. For example, there exists a polynomial $p \in \mathbb{N}[x]$, so that $\|A(BC) = (AB)C\|_n$ has Frege proofs of size $\leq p(n)$.*

See [6] for a proof of this theorem.

# 3 eFrege and Gaussian Elimination

In this section we show that the correctness of Gaussian Elimination can be proven with short eFrege proofs, and we show that because of that, hard matrix identities also have short eFrege proofs.

## 3.1 Correctness of GE

Recall that a matrix is in *row-echelon form* if it satisfies the following two conditions: (i) if there is a non-zero row, the first non-zero entry of every row is 1, (the *pivot*), and (ii) the first non-zero entry of row $i+1$ is to the right of the first non-zero entry of row $i$. In short, a matrix is in row-echelon form if it looks as follows:

$$\left[\begin{array}{cccccccc} 1 & * \dots * & * & * \dots * & * & * \dots * & * & \\ & & 1 & * \dots * & * & * \dots * & * & \\ & \ddots & & & 1 & * \dots * & * & \\ & & 0 & & & & 1 & \dots \\ & & & \ddots & & & \vdots & \ddots \end{array}\right] \tag{6}$$

where the $*$'s indicate entries from $\mathbb{Z}_2$.

We define the function Gaussian Elimination, $GE : M_{n \times m} \longrightarrow M_{n \times n}$, to be the function which given an $n \times m$ matrix $A$ as input, it outputs an $n \times n$ matrix

$GE(A)$, with the property that $GE(A)A$ is in row-echelon form. We call this property the *correctness condition* of $GE$.

We show how to compute $GE(A)$, given $A$. The idea is, of course, that $GE(A)$ is equal to a product of elementary matrices which bring $A$ to row-echelon form. We start by defining elementary matrices. Let $T_{ij}$ be a matrix with zeros everywhere except in the $(i, j)$-th position, where it has a 1. A matrix $E$ is an *elementary matrix* if $E$ has one of the following three forms:

$$\begin{aligned} &I + aT_{ij} \quad i \neq j && \text{(elementary of type 1)} \\ &I + T_{ij} + T_{ji} - T_{ii} - T_{jj} && \text{(elementary of type 2)} \\ &I + (c-1)T_{ii} \quad c \neq 0 && \text{(elementary of type 3)} \end{aligned}$$

Let $A$ be any matrix. If $E$ is an elementary matrix of type 1, then $EA$ is $A$ with the $i$-th row replaced by the sum of the $i$-th row and $a$ times the $j$-th row. If $E$ is an elementary matrix of type 2, then $EA$ is $A$ with the $i$-th and $j$-th rows interchanged. If $E$ is an elementary matrix of type 3, then $EA$ is $A$ with the $i$-th row multiplied by $c \neq 0$.

We compute $GE$ recursively, on the number of rows of $A$. If $A$ is a $1 \times m$ matrix, $A = [a_{11}a_{12}\dots a_{1m}]$, then:

$$GE(A) = \begin{cases} [1/a_{1i}] & \text{where } i = \min\{1, 2, \dots, m\} \text{ such that } a_{i1} \neq 0 \\ [1] & \text{if } a_{11} = a_{12} = \dots = a_{1m} = 0 \end{cases} \tag{7}$$

In the first case, $GE(A) = [1/a_{1i}]$, $GE(A)$ is just an elementary matrix of size $1 \times 1$, and type 3, $c = a_{i1}$. In the second case, $GE(A)$ is a $1 \times 1$ identity, so an elementary matrix of type 1 with $a = 0$. Also note that in the first case we divide by $a_{1i}$. This is not needed when the underlying field is $\mathbb{Z}_2$, since a non-zero entry is necessarily 1. However, we claim throughout this paper that our arguments hold regardless of the underlying field, so we want to make the function GE field independent.

Suppose now that $n > 1$. If $A = 0$, let $GE(A) = I$. Otherwise, let:

$$GE(A) = \begin{bmatrix} 1 & 0 \\ 0 & GE((EA)[1|1]) \end{bmatrix} E \tag{8}$$

where $E$ is a product of at most $n+1$ elementary matrices, defined below. Note that $C[i|j]$ denotes the matrix $C$ with row $i$ and $j$ deleted, so $(EA)[1|1]$ is the matrix $A$ multiplied by $E$ on the left, and then the first row and column are deleted from the result. Also note that we make sure that $GE(A)$ is of the appropriate size (i.e., it is an $n \times n$ matrix), by placing $GE((EA)[1|1])$ inside a matrix padded with a 1 in the upper-left corner, and zeros in the remaining of the first row and column.

**Definition of $E$:** If the first column of $A$ is zero, let $j$ be the first non-zero column of $A$ (such a column exists by the assumption $A \neq 0$). Let $i$ be the index of the first row of $A$ such that $A_{ij} \neq 0$. If $i > 1$, let $E = I_{1i}$ ($E$ interchanges row 1 and row $i$). If $i = 1$, but $A_{lj} = 0$ for $1 < l \leq n$, then $E = I$ (do nothing). If $i = 1$, and $1 < i'_1 < i'_2 < \cdots < i'_k$ are the indices of the other rows with $A_{i'_l j} \neq 0$, let $E = E_{i'_1} E_{i'_2} \cdots E_{i'_k}$, where $E_{i'_l}$ is the elementary matrix that adds the first row of $A$ to the $i'_l$-th row, of $A$ so that it clears the $j$-th entry of the $i'_l$-th row (this is over $\mathbb{Z}_2$; over a bigger field, we might need *a multiple* of the first row to clear the $i'_l$-th row).

If the first column of $A$ is not zero, then let $a_{i1}$ be its first non-zero entry (i.e., $a_{j1} = 0$ if $j < i$). We want to compute a sequence of elementary matrices, whose product will be denoted by $E$, which accomplish the following sequence of steps:

1. they interchange the first and $i$-th row,
2. they divide the first row by $a_{i1}$,
3. and they use the first row to clear all the other entries in the first column.

Let $a_{i_1 1}, a_{i_2 1}, \ldots, a_{i_k 1}$ be the list of all the non-zero entries in the first column of $A$, not including $a_{i1}$, ordered so that:

$$i < i_1 < i_2 < \cdots < i_k$$

Let the convention be that if $a_{i1}$ is the *only* non-zero entry in the first row, then $k = 0$. Define $E$ to be:

$$E = E_{i_1} E_{i_2} \cdots E_{i_k} E' E''$$

where $E_{i_j} = I - a_{i_j 1} T_{i_j 1}$, so $E_{i_j}$ clears the first entry from the $i_j$-th row of $A$. Note that if $k = 0$ (if $a_{i1}$ is the only non-zero entry in the first column of $A$), then $E = E''E'$. Let

$$E'' = I + \left( \frac{1}{a_{i1}} - 1 \right) T_{11} \quad \text{and} \quad E' = I + T_{i1} + T_{1i} - T_{ii} - T_{11}$$

Thus, $E''$ divides the first row by $a_{i1}$, and $E'$ interchanges the first row and the $i$-th row. **End of definition of $E$.**

We define the Boolean formula $\mathit{RowEchelon}(C_{11}, C_{12}, \ldots, C_{nm})$ to be the disjunction of (9) and (10) below:

$$\bigwedge_{1 \leq i \leq n, 1 \leq j \leq m} \neg C_{ij} \tag{9}$$

$$\bigwedge_{1 \leq i < n, 1 < j \leq m} \left( (\neg C_{(i+1)1} \wedge \ldots \wedge \neg C_{(i+1)(j-1)} \wedge C_{(i+1)j}) \supset \bigvee_{1 \leq k \leq j-1} C_{ik} \right) \tag{10}$$

Note that (9) states that $C$ is the zero matrix, and (10) states that the first non-zero entry of row $i+1$ is to the right of the first non-zero entry of row $i$. Moreover, if the $(i+1)$-st row has a non-zero entry, then the $i$-th row *must* also have a non-zero entry. Note that we do not need to state the condition that the first non-zero entry of each row is 1, since the field is $\mathbb{Z}_2$; over more general fields, we would have to state this condition as well.

We will abuse notation slightly, and sometimes write $\mathit{RowEchelon}(C)$ in place of $\mathit{RowEchelon}(C_{11}, C_{12}, \ldots, C_{nm})$.

**Theorem 2** *eFrege proves the correctness of GE with proofs of size polynomial in the given matrix. More precisely, the family of tautologies given by:*

$$\{\bigwedge \|C = GE(A)A\|_{n,m} \supset \mathit{RowEchelon}(C)\} \tag{11}$$

*has short eFrege proofs.*

Proof: We prove that (11) has short eFrege proofs. More precisely, from the constructions of the derivations given below, it is possible to come up with a constant $d$, so that the size of these derivations (measured in the number of symbols) is bounded by $(n+m)^d$, $n, m \geq 1$. We do not give $d$ explicitly.

We build the proof of (11) inductively on $n$. Suppose first that $A$ is a $1 \times m$ matrix. Let $G = GE(A)$, then from (7) we see that $G = [1]$, so it is represented by the single extension definition $G_{11} \leftrightarrow 1$. Now, define $C = GA$ with $m$ extension definitions, and show that $\bigwedge \|C = A\|_{1,m}$. Since $A$ has only one row, and it is a matrix over $\mathbb{Z}_2$, it follows that $A$ is in row-echelon form, and hence $\mathit{RowEchelon}(C)$ follows.

Now suppose that $A$ is a $(n+1) \times m$ matrix. Let $G' = GE((EA)[1|1])$, and we already have the set of extension definitions for $G'$ by induction. Thus, from:

$$G = \begin{bmatrix} 1 & 0 \\ 0 & G' \end{bmatrix} E$$

we obtain the set of extension definitions for $G = GE(A)$. This set is short because the definition of $E$ is short (see (3) for the definition of iterated matrix products), and because the definition of $G'$ is short, by induction. More precisely, $E$ is given by at most $n+2$ elementary matrices of size $(n+1) \times (n+1)$ each; thus, it involves $n+1$ new matrix definitions, each definition of size bounded by $O((n+1)^3)$ (just recall the definition of $\|C = AB\|_{n+1}$). Each of the elementary matrices that make up $E$ (see the definition of $E$ above), over $\mathbb{Z}_2$, has a definition of constant size (in terms of the entries of $A$). Thus, the extension definitions of $E$ are of size bounded by $O((n+1)^4)$. Therefore, $G$ can be defined with $O((n+1)^4)$ + (nr. of extension definitions for $G'$) extension definitions, which is $O(\sum_{k=1}^{n+1} k^4) \leq O((n+1)^5)$ many extension definitions in total for $G$.

Let $C' = G'((EA)[1|1])$, and $C = GA$. By induction,

$$\bigwedge \|C' = G'((EA)[1|1])\|_n \supset \mathit{RowEchelon}(C')$$

has an eFrege proof of size bounded by $(n+m)^d$. We now want to show that given the extension definitions for $G'$ and $G$, $\mathit{RowEchelon}(C') \supset \mathit{RowEchelon}(C)$ has short eFrege proofs. Since

$$C = GA = \begin{bmatrix} 1 & 0 \\ 0 & G' \end{bmatrix} EA = \begin{bmatrix} \multicolumn{2}{c}{\text{first row of } EA} \\ 0 & G'((EA)[1|1]) \end{bmatrix} = \begin{bmatrix} \text{first row of } EA \\ 0 \quad C' \end{bmatrix}$$

To see this, note that the first column of $EA$ is zero, except possibly for the first entry. By the choice of $E$, either $(EA)_{11} \neq 0$, in which case we have $RowEchelon(C)$, or the first non-zero entry of the first row of $EA$ is to the left of the first non-zero column of $C'$, in which case we also have $RowEchelon(C)$. Also note that we use associativity of iterated matrix products in the above reasoning. That is, we assume that the way we parenthesize an iterated matrix product is not important, since by associativity we always get the same result. This can be shown with short eFrege proofs as well. □

**Theorem 3** *The existence of the inverse of $GE(A)$ can be shown with short eFrege proofs.*

Proof: We have to show that given $\|G = GE(A)\|_n$, the Boolean variables $G^{-1}_{11}, G^{-1}_{12}, \ldots, G^{-1}_{nn}$, corresponding to $G^{-1}$, can be constructed with short extension definitions, and that eFrege proves $\|GG^{-1} = I\|_n$ with short proofs.

Just as we defined $G$ inductively with extension definitions, we define $G^{-1}$ inductively. Given $E = E_{i_1}E_{i_2}\cdots E_{i_k}E'E''$, we can compute $E^{-1}$ immediately by letting it be $E''^{-1}E'^{-1}E^{-1}_{i_k}\cdots E^{-1}_{i_2}E^{-1}_{i_1}$. Each of these inverses can be computed very easily, because they are elementary matrices. So, since we are dealing with $\mathbb{Z}_2$, $E'' = E''$, and $E'$ is also its own inverse, and $E_{i_j}$ is a matrix with 1s on the diagonal, and 1 in the position $(p, q)$, so $E^{-1}_{i_j}$ is a matrix with 1s on the diagonal, and a 1 in position $(q, p)$.

Thus, we showed how to compute $G^{-1}$. We still need to show that the family of tautologies $\{\|GG^{-1} = I\|_{n,m}\}$ has short eFrege proofs, for any $n \times m$ matrix $A$. We can prove this inductively on the number of rows of $A$, just as in the proof of Theorem 2, so we do not repeat it here. □

**Corollary 1** *It can be shown with short eFrege proofs that $GE(A)A$ has 1s on the main diagonal, or its last row is zero.*

Proof: The truth of this assertion is obvious from (6). Let $C = GA$, and suppose that there is a zero entry on the diagonal, i.e., $\neg \bigwedge_{1\leq i\leq n} C_{ii} \leftrightarrow 1$. We want to show that the last row is zero, $\bigwedge_{1\leq i\leq n} C_{ni} \leftrightarrow 0$. We know that $RowEchelon(C)$ is valid, and provable in polysize eFrege (by Theorem (2)). From (10) we can conclude with short eFrege proofs that:

$$\bigwedge_{1\leq j\leq k} \neg C_{ij} \supset \bigwedge_{1\leq j\leq k+1} \neg C_{(i+1)j} \tag{12}$$

That is, if the first $k$ entries of row $i$ are zero, then the first $(k+1)$ entries of row $(i+1)$ are zero. Let $C_{ii}$ be the zero, with the smallest $i$. Now, from (12) we prove that:

$$\bigwedge_{1\leq j\leq i} C_{ij} \leftrightarrow 0 \tag{13}$$

Using (12) repeatedly, for $0 \leq k \leq n - i$, we show that the first $(i+k)$ entries of row $(i+k)$ are zero. Thus, we can conclude that the first $n$ entries of the $n$-th row are zero, and, therefore, the $n$-th (last) row is zero altogether.

In fact, note that given $RowEchelon(C)$, all we needed was polysize Frege to prove that if some $C_{ii}$ is zero, then the last row of $C$ is zero. □

### 3.2 Extended Frege proofs of hard matrix identities

It was shown in [6], that the identities I, II, III, and IV, presented in section 2.1, are equivalent, and furthermore, their equivalence can be shown with short Frege proofs. Thus, by showing that one of the four identities has short eFrege proofs, we show that they all do. We choose $AB = I \supset BA = I$.

**Theorem 4** *The tautologies $\|AB = I \supset BA = I\|_n$ have short eFrege proofs.*

Proof: Suppose that $AB = I$. By Corollary 1, $GE(A)A$ has 1s on the diagonal, or its bottom row is zero. Since $AB = I$, it follows that:

$$GE(A) = GE(A)(AB) = (GE(A)A)B$$

so, if the bottom row of the matrix $GE(A)A$ is zero, then so is the bottom row of the matrix $(GE(A)A)B$, and hence so is the bottom row of $GE(A)$, which by Theorem 3 is not possible (since if a matrix has a row of zeros, it cannot have an inverse). Hence, $GE(A)A$ has 1s on the diagonal. Using a sequence of elementary matrices $F_1 \ldots F_k$, we can clear the entries above the main diagonal, so that:

$$F_k F_{k-1} \cdots F_1 GE(A)A = I$$

and hence $C = F_k F_{k-1} \cdots F_1 GE(A)$ is the left inverse of $A$.

Now that we know that $A$ has a left inverse, and since we can show with basic ring properties that $AB = I \supset A(BA - I) = I$, it follows that $BA = I$. To see that $AB = I \supset A(BA - I)$ can be proven with ring properties, note that if $AB = I$, then $(AB)A = A$, so by associativity, $A(BA) = A$, so $A(BA) - A = 0$, so by left-distributivity, $A(BA - I) = 0$. Also, since $AB = I \supset A(BA - I) = 0$ can be shown with ring properties, it follows that it can be proven with polysize Frege proofs. □

## 4 Conclusions and Open Problems

We have seen that the correctness of the Gaussian Elimination algorithm can be proven with polysize eFrege proofs. That is, the tautologies which express the relation "$GE(A)A$ is in row-echelon form," have uniform polysize (in size of $A$) eFrege proofs. Our proofs work over the field $\mathbb{Z}_2$, but all the results can be replicated for bigger fields, such as $\mathbb{Z}_p$ or $\mathbb{Q}$.

We have also seen that because Gaussian Elimination has short eFrege proofs of correctness, the matrix identity $AB = I \supset BA = I$, and hence a host of matrix identities (we called them "hard matrix identities"), have short eFrege proofs. In fact, many combinatorial principles, such as the "Odd Town Theorem," can also be proven with short eFrege proofs; see [6] for details.

Two important questions arise, for future research: (i) are "hard matrix identities" indeed *hard* ? The identity $AB = I \supset BA = I$ has been proposed by Cook as a candidate for separating Frege and eFrege; we have seen that it has short eFrege proofs, so if it is hard for Frege, we would have the separation of Frege and eFrege. If it turns out that $AB = I \supset BA = I$ has short Frege proofs after all, then we could take it as evidence that we have to work harder on trying to prove that Frege is just as strong as eFrege, besides trying to prove the separation.

(ii) Can "hard matrix identities" be proven in quasi-polysize Frege ? In [6], it is shown how to construct a theory of linear algebra based on Berkowitz's algorithm (which computes the characteristic polynomial of a matrix), rather than on the Gaussian Elimination algorithm. The advantage of this approach is that, while Gaussian Elimination is a sequential polytime algorithm, Berkowitz's algorithm is a fast parallel algorithm, that runs in time $O(\log^2 n)$, where $n$ is the size of the given matrix. Berkowitz's algorithm can be formalized with quasi-polysize Boolean formulas, but it is not known how to prove it correct with such formulas. If we could prove it correct with short quasi-polysize Frege, we would also have short quasi-polysize Frege of $AB = I \supset BA = I$. This would be an important step forward in understanding the complexity of proofs of matrix identities.

**Acknowledgments:** This research originates from the author's Ph.D. thesis, written under the supervision of Stephen Cook (at the University of Toronto). Zbigniew Stachniak (York University) read the original manuscript, and gave valuable comments.

# References

[1] Miklós Ajtai. The complexity of the pigeonhole principle. In *29th Annual Symposium on Foundations of Computer Science*, pages 346–355. IEEE, 1988.

[2] Paul Beame and Toniann Pitassi. Propositional proof complexity: Past, present, and future. *Bulletin of the EATACS, TR98-067*, 1998.

[3] Stephen A. Cook and Robert A. Reckhow. The relative efficiency of propositional proof systems. *JSL*, 44:36–50, 1979.

[4] A. Haken. The intractability of resolution. *Theoretical Computer Science*, 39:297–305, 1985.

[5] Jan Krajíček. *Bounded Arithmetic, Propositional Logic, and Complexity Theory.* Cambridge, 1995.

[6] Michael Soltys. *The complexity of derivations of matrix identites.* PhD thesis, University of Toronto, 2001.

[7] Michael Soltys. Berkowitz's algorithm and clow sequences. *Electronic Journal of Linear Algebra*, 9:42–54, 2002.

[8] Michael Soltys and Stephen Cook. The proof complexity of linear algebra. In *Symposium on Logic in Computer Science (LICS'2002)*. IEEE, 2002.

[9] Alasdair Urquhart. The complexity of propositional proofs. *Bulletin of Symbolic Logic*, 1(4):425–467, 1995.

[10] Joachim von zur Gathen. Boolean circuits versus arithmetic circuits. *Information and Computation*, 91:142–154, 1991.