

דב פרוהמן — ממציא הזיכרון הבלתי-נדיף

Non-Volatile Memory Eprom Dov Frohman — Inventor

מעבודת הדוקטורט ועד הקמת אינטל ישראל — מ-EPROM ל-SSD היום

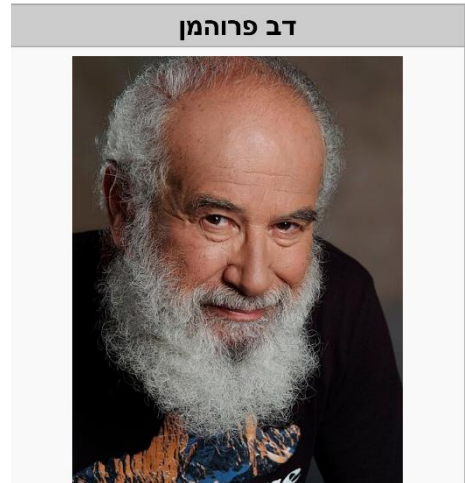
From PhD Thesis to Intel Israel — From EPROM to Today's SSD

מחקר ועריכה- איסי גייר, מיסמך לא רישמי של אינטל יעוץ, אהוד קפלן

המאמר מכיל גם רשמים אישיים של איסי גייר מ Fab8



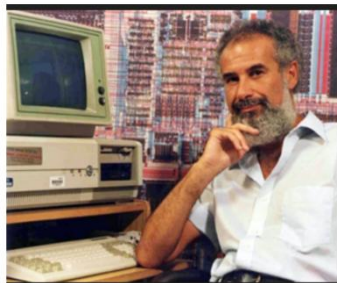
ד"ר דב פרוהמן, אינטל מ 1971 ממציא ה-EPROM ומייסד אינטל ישראל (Computer History Museum מדגים את EPROMה)



דב פרוהמן



דב פרוהמן EPROM



דב פרוהמן מעבר לויקיפדיה <

סיכום:

מסמך זה מציג ניתוח מקיף של אחד ממסמכי היסוד החשובים ביותר בהיסטוריה של תעשיית המוליכים למחצה: "Charge Transport and Storage in Metal-Nitride-Oxide-Silicon (MNOS) Structures" — מאמר שפרסם ד"ר דב פרוהמן-בנצ'קובסקי עם M. Lenzlinger בשנת 1969, ואשר הניח את הבסיס הפיזיקלי לכל זיכרון לא-נדיף מאז ועד היום.

נקודות מפתח

הבעיה שנפתרה	כיצד לשמר מידע בתוך רכיב MOS ללא מקור מתח רציף — הבסיס לזיכרון לא-נדיף
הפתרון הפיזיקלי	לכידת מטען (charge trapping) בממשק SiO_2/Si_3N_4
מנגנון כתיבה	Fowler-Nordheim tunneling דרך תחמוצת דקה; Poole-Frenkel בניטרוד

המשמעות ההיסטורית	פרוהמן התחיל את הרעיון ל-EPROM 1702 (1971) ← Floating Gate ← Flash ← SSD
ההשפעה התעשייתית	EPROM היה המוצר הרווחי ביותר של אינטל לשנות ה-80; כל Flash ו-SSD כיום בנוי על אותה פיזיקה
Intel Israel	פרוהמן ייסד את Intel Israel (1974), ממרכזי ה-R&D המובילים בעולם, ב-1982 יסד את מפעל הסיליקון FAB 8 בירושלים.
קשרהשתלת יונים	אותה פיזיקת trapping בדיאלקטרי רלוונטית להבנת charging damage בתהליכי השתלת יונים

גורדון מור, מייסד-שותף של אינטל, על ה-EPROM: "חשוב לא פחות מהמיקרופרוססור עצמו בהתפתחות ענף המחשבים האישיים"

תקציר

מאמר זה מנתח לעומק את עבודתם של **M. Lenzlinger** ו-**Dov Frohman-Bentchkowsky** משנת 1969, שפורסמה ב-*Journal of Applied Physics*, תחת הכותרת "*Charge Transport and Storage in Metal-Nitride-Oxide-Silicon (MNOS) Structures*". עבודה זו מהווה ציון דרך מרכזי בהיסטוריה של מוליכים למחצה: היא מניחה את הבסיס הפיזיקלי לזיכרון לא-נדיף בטכנולוגיית MOS, ומובילה ישירות להמצאת ה-EPROM ב-1971 ועד לדיסק און קי וכונן SSD

פרק א': דב פרוהמן באינטל: סיפור ההמצאה של ה-EPROM

1.1 הרקע האישי: מאמסטרדם לסיליקון ואלי

דב פרוהמן נולד במרץ 1939 באמסטרדם, חמישה חודשים לפני פרוץ מלחמת העולם השנייה. הוריו, אברהם ופייגה פרוהמן, יהודים פולנים שהיגרו להולנד בשנות ה-30. ב-1942, עם הידוק האחיזה הנאצית, הפקידו הוריו את הילד הקטן בידי משפחת **ואן טילבורך** — משפחה נוצרית-קלוויניסטית בכפר ספרנג-קאפל שבדרום הולנד — שהסתירה אותו לאורך שנות הכיבוש. הוריו לא שרדו את השואה.

לאחר המלחמה גדל פרוהמן בישראל, שם למד הנדסת חשמל בטכניון. לאחר שירות צבאי בגולני ב-עבר לארצות הברית ללימודי תואר שני ודוקטורט. תוך כדי הלימודים עבד במעבדות

Fairchild Semiconductor בעמק הסיליקון — שם פגש בצוות המיתולוגי שיקים את אינטל: גורדון מור, רוברט נויס, אנדי גרוב ב-1969, עם סיום הדוקטורט בברקלי על MNOS, גייס אותו **אנדי**

גרוב לאינטל. פרוהמן הצטרף לחברה הצעירה — שהיתה כבת שנה — ולקח על עצמו עבודה בפיתוח זיכרוני MOS.

1.2 SRAM 1101 — נקודת המוצא: תקלת אמינות

בסתיו 1969, אינטל חוותה בעיות אמינות חמורות בשבב

Intel 1101 — ה-SRAM הסטטי הראשון שיוצר בתהליך *silicon-gate MOS*. זה היה שבב חדשני: במקום שער מתכת רגיל, השתמשו בפוליסיליקון כחומר השער — מה שאפשר תהליך ייצור מדויק יותר. אולם בשטח, 1101 הפגין התנהגות לא יציבה. מתח הסף של חלק מהטרנזיסטורים השתנה לאורך זמן. **פרוהמן הוטל לחקור את הגורם.**

מה גילה? שבמצבים מסוימים, שכבת SiO_2 — הבידוד של שער הטרנזיסטור — **סופגת אלקטרונים** מרכיבי המתכת הסמוכים ומחזיקה מטען חשמלי. המטען הזה מפריע לתפקוד הרגיל ומזיז את מתח הסף.

בשפת ה-MNOS שפרוהמן הכיר היטב מהדוקטורט: זה היה **charge trapping** — לכידת מטען בדיאלקטרי. רק שכאן זה קרה ב-SRAM, לא בזיכרון מכונן.

1.3 הרגע שבין תיקון להמצאה

רוב המהנדסים היו מטפלים בבעיה בצורה ישירה: למנוע את ה-trapping ולייצב את הרכיב. פרוהמן עשה זאת — אבל גם שאל את השאלה הפוכה:

"ואם אני אשתמש בזה?"

אם אפשר להגרם ולשלוט בלכידת המטען ב- SiO_2 — אפשר ליצור שבב שניתן לתכנות חשמלית, ישמר את המידע ללא מתח, ויאפשר איפוס. כלומר:

זיכרון שניתן למחיקה ולכתיבה חוזרת.

פרוהמן עצמו תיאר בראיון ל-Computer History Museum (2009) שאין **"רגע האאהא"** יחיד:

"אני לא חושב שרגע האאהא קיים. אני לא יודע מה היה התובנה המרכזית, אבל ברגע שהיתה לי אותה, שאר התהליך היה כמעט צפוי."

1.4 הפיזיקה של FAMOS — מה שפרוהמן בנה

הפתרון שפרוהמן פיתח נקרא

FAMOS — Floating-Gate Avalanche-injection MOS. זהו טרנזיסטור MOS שבו השער הוא **פוליסיליקון מבודד לחלוטין** — אין לו חיבור חשמלי לשום חלק אחר בשבב. הוא "צף" בתוך SiO_2 .

מבנה FAMOS:

- ערוץ p-type בסיליקון
- תחמוצת SiO₂ מעל הערוץ
- שער צף (floating gate) — פוליסייליקון מבודד מכל עבריו
- תחמוצת נוספת מעל השער הצף
- אין שכבת שליטה עליונה (בניגוד ל-EEPROM מאוחר יותר)

כתיבה — Avalanche Injection: מפעילים מתח גבוה (~40 וולט) בין המקור לניקוז. השדה החזק גורם ל**breakdown מקומי בצומת n-p** שמתחת לשער — תהליך המכונה avalanche breakdown. ה"אלקטרונים החמים" (hot electrons) שנוצרים מואצים כל כך שיש להם מספיק אנרגיה לחצות את שכבת SiO₂ ולהיקלט על השער הצף.

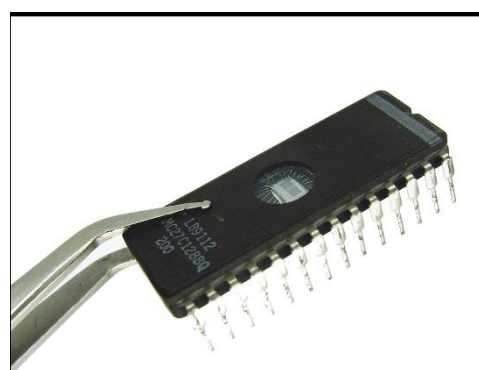
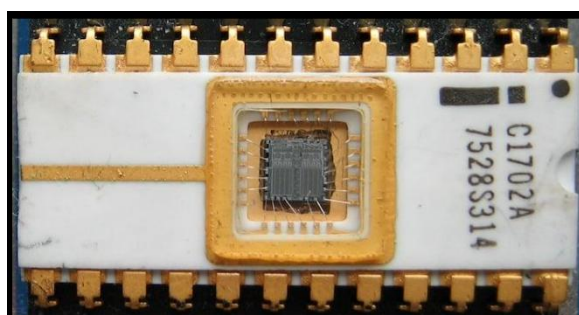
ברגע שהמטען על השער הצף — הוא **נשאר שם**. אין לו לאן ללכת: מכל צד תחמוצת. המטען משנה את מתח הסף של הטרנזיסטור — ניתן לקרוא זאת.

מחיקה — קרינת UV: כדי לאפס, מאירים את השבב בקרינת אולטרה-סגול (~253.7 ננומטר, אנרגיית פוטון ~4.9 eV). האנרגיה הזאת מספיקה כדי להעניק לאלקטרונים בשער הצף מספיק אנרגיה לחמוק דרך ה-SiO₂ חזרה לסיליקון. תהליך המחיקה לוקח

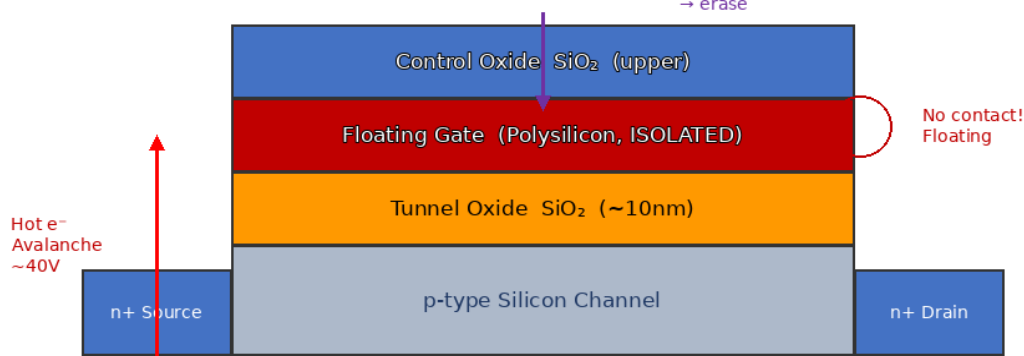
10–30 דקות בנורת UV סטנדרטית.

כדי לאפשר כניסת UV, חבילת ה-EPROM כללה

חלון קוורץ שקוף מעל השבב — מה שהפך את ה-EPROM לניתן לזיהוי מיידי לכל מהנדס שראה אותו.



FAMOS / EPROM Floating Gate Structure (Frohman 1971)



Write: ~40V avalanche injects hot e⁻ onto floating gate | Erase: UV photons (~4.9eV) release charge

איור 2 — מבנה FAMOS/EPROM: Floating Gate (Frohman-Bentchkowsky, הזרקה אלקטרונים חמים ומחיקת UV), *Applied Physics Letters*, 1971

1.5 האתגרים בדרך לפרוטוטיפ

הדרך מהרעיון לפרוטוטיפ היתה לא היתה פשוטה. פרוהמן תיאר בראיון:

הספקנות הפנימית באינטל: כשהציג את הרעיון בפני גורדון מור, רוברט נויס, אנדי גרוב ולס ואדז — "היה הרבה ספקנות". הרעיון של שמירת מטען על שער צף ללא כל חיבור נחשב ל"רעיון מטורף". ב-MNOS ובמבנים דומים, תמיד היה שני שערים — שניתן להבין את המנגנון. ה-FAMOS עם שער צף בלתי מחובר נראה כמעט מאגי.

הבעיה של ה-retention: כדי להוכיח שהמטען יישמר, פרוהמן הניח EPROM's על גג בניין אינטל ובדק שאור שמש רגיל אינו מוחק אותם. ימים ארוכים חלפו — המידע נשמר. לאחר בדיקות ממושכות הוא הצליח לשכנע את הנהלת אינטל שה-retention עומד בשנים רבות בתנאים רגילים.

הבעיה של ה-avalanche injection: מנגנון ה-avalanche כולל תהליכים מורכבים — coupling קיבולי, הולכה בשדה גבוה, ספיגה בתחמוצת. פרוהמן עצמו הודה שהמנגנון המדויק היה

"הרבה יותר מורכב ממה שחשבתי בהתחלה".

עבודה סולו: פרוהמן עבד בהתחלה לבד לחלוטין. הפרויקט לא היה מאורגן, לא היה לו תקציב רשמי. הוא התייעץ עם לס ואדז בנקודות מפתח, אבל העבודה הניסויית נעשתה על ידו בלבד.

1.6 הכנס שבו הכל השתנה: ISSCC 1971

בפברואר 1971 הציג פרוהמן את ה-FAMOS בכנס ISSCC — International Solid-State Circuits Conference בפילדלפיה.

ההדגמה היתה דרמטית: פרוהמן הציג סרט שהראה את תבנית הביטים ב-EPROM — ביטים שיצרו את **לוגו Intel**. אחר כך חשף את השבב לאור UV. הביטים נפלו אחד אחרי השני — עד שנמחק לוגו אינטל לחלוטין. גורדון מור תיאר את הרגע: "הביטים נפלו, וכשנפל האחרון, כל הקהל פרץ מחיאות כפיים." מאמרו של פרוהמן נבחר להמאמר הטוב ביותר בכנס.

ב-1971 אינטל שחררה את ה-**Intel 1702** — ה-EPROM המסחרי הראשון בעולם. קיבולת: 2048 ביט (256 בייט). מחיר: גבוה. שוק: בינתיים — נישה.



Fellow Award Honoree 2018 וידאו מטקס הענקת הוקרה לדוב Introduction & Remarks—Dov Frohman-Bentchkowsky

אם הקישור לא נפתח בטלפון, להעתיק את הקישור לגוגול כרום ולפתוח

<https://www.youtube.com/watch?v=kUhdYKdz-Mg>

1.7 המפגש עם המיקרופרוססור — הצמד המכונן

ב-1971 אינטל פרסמה גם את **Intel 4004** — המיקרופרוססור המסחרי הראשון בעולם. ברגע זה נולד הצמד המושלם:

- המיקרופרוססור דורש תוכנה (firmware, microcode).
 - התוכנה חייבת להישמר גם ללא מתח — זיכרון לא-נדיף.
 - ROM מסורתית דרש ייצור מסכה מחדש (שבועות ואלפי דולרים) לכל שינוי.
 - EPROM אפשר לתכנת ולמחוק תוך שעות.
- פרוהמן עצמו הסביר שבהתחלה ראו ב-EPROM רק "כלי אבטיפוס" — מכשיר לפיתוח, לא לייצור סדרתי. אבל המיקרופרוססור שינה את המשוואה: "*EPROM allowed engineers to iterate*" — "firmware and logic designs without costly mask changes".
- גורדון מור הגדיר את ה-EPROM כ-

"חשוב לא פחות מהמיקרופרוססור עצמו בהתפתחות ענף המחשבים האישיים." זמן הפיתוח של אבטיפוסים חדשים ירד מ שבועות לשעות. זה שינה את תעשיית הסמיקונדוקטורים לעולמים.

1.8 מאמרי FAMOS — הפרסומים המרכזיים

חשיבות	פרסום	שנה
--------	-------	-----

אפריל 1971	,Applied Physics Letters Vol.18(8), pp.332-334	FAMOS — מאמר פיזיקלי ראשון
אוקטובר 1971	,IEEE JSSC Vol.SC-6(5) pp.301-306	FAMOS ROM-2048 bit מלא
1972	,Solid-State Electronics Vol.17, pp.517-529	FAMOS — מאפיינים מלאים
1980	IEEE Trans. Electron ,Devices ED-27(9) pp.1744-1752	מודל מחיקה ל-FAMOS (עם קצנלסון)

1.10 פרוהמן עוזב אינטל — ועובד בגאנה

ב-1972, שנה לאחר הצגת ה-EPROM, עשה פרוהמן צעד שהפתיע את כולם: עזב את אינטל ועבר כפרופסור אורח לאוניברסיטת המדע והטכנולוגיה בקומאסי, גאנה. מטרתו: נסיעה ולמידה, תמיכה בפיתוח חינוך הנדסי באפריקה. ב-1974 חזר לישראל, הצטרף לאוניברסיטה העברית בירושלים כפרופסור עמית בפיזיקה יישומית, ובמקביל ייעץ לאינטל — עד שהוביל אותם לצעד ההיסטורי הבא: הקמת Intel Israel.

1.11 Intel Israel — מרכז חיפה ועד Fab-8 בירושלים

ב-1974 פרוהמן הקים את מרכז תכנון שבבים קטן בחיפה — מרכז ה-R&D הראשון של אינטל מחוץ לארצות הברית. ממליץ לראות את הראיון עם רפי נווה על הקמת אינטל חיפה וההצלחה הגדולה <https://www.youtube.com/watch?v=JRSUmuWiTOs>

. ב-1981 הפך למנהל הכללי של Intel Israel. בעיקבות ההצלחה של אינטל חיפה, דוב יזם הקמת מיפעל ליצור סמיקונדוקטורס, בירושלים, דוב גר בעין כרם בירושלים. פרוהמן ניהל משא ומתן ממושך עם הממשלה הישראלית החל מ-1979, עד שב-1981 יצאה לדרך

Fab-8 בירושלים — מפעל ייצור שבבי סיליקון הראשון של אינטל מחוץ לארצות הברית. זה היה רגע היסטורי לתעשיית ההיי-טק הישראלית.

בנובמבר 1982 דוב ריאיון אותי לאינטל, ללמוד ולהביא את טכנולוגית משתילי יונים.

זיכרון אישי

ראיון עם דב פרוהמן

הראיון

בנובמבר 1982 עברתי ראיון אצל דב פרוהמן לקראת קבלה לעבודה באינטל. זה היה, ראיון שלישי.

באותה תקופה לימדתי אלקטרוניקה מיקרו-מחשבים במיכללה להנדסאים בבית-ספר טכנולוגי, בכיתות י"ג-י"ד.

בנוסף להוראת אלקטרוניקה הינחתי פרויקטי גמר בתחום מחשבים מיקרו-מעבדיים, והדרכת סטודנטי בתכנון ובביצוע מערכות מבוססות מיקרו-מחשב. במקביל עבדתי בחברת מומחי מחשב לבנות מערכות ללהבי ישקר למערכות CNC. ולבית חולים שערי צדק מחלקה נירולוגית.

במהלך הראיון שאל אותי דב מספר שאלות בתחום האלקטרוניקה.

בסיום הדיון אמר לי שעברתי את הראיון.

שאלתי אותו:

"מה תהיה עבודתי?"

דב ענה:

העבודה תותאם לכישורך

כבדיחה אפשר לומר היום שתשובה זו לא נתנה לי הבנה ממשית של מה שעמדתי לעשות בפועל.

לאחר מכן שלף דב חוברת עם תמונה מכונת השתלת יונים של וריאן DF4 מיפלצתית ואמר:

"אתה תלמד את זה."

באותו רגע הנחתי שכנראה אעסוק בעיקר באלקטרוניקה.

מה שקרה במציאות היה שונה לחלוטין ומרתק. הרבה יותר.

ראשית הדרך בייצור מוליכים למחצה

הראיון עם דב פרוהמן סימן את תחילת מעורבותי בייצור מוליכים למחצה.

כשהגעתי בינואר 83 לפאב-5 לצורך הכשרה, לקחו אותי אל חדר הנקי והחלו להסביר לי את הסביבה שבה אעבוד.

סיפרו לי שמשתיל היוניים כולל מתחים של עד 200,000 וולט, גזי תהליך שונים מסוכנים ארסין, פוספין, בורון מערכות ואקום גבוה עם משאבות דיפוזיה ושמן רותח, חנקן נוזלי, ומשאבות ואקום מכניות מלאות בשמן מזוהם המכילים חומרים מסוכנים.

עבורי זה היה זעזוע של ממש.

בתמימות שאלתי:

"היכן מנתקים את הכוח מהמכונה כשמבצעים תחזוקה?"

הם הסתכלו עליי וענו משהו בסגנון:

"אתה משוגע? אם ננתק את הכוח מהמכונה יידרשו כ-12 שעות להגיע שוב לוואקום גבוה."

עולם אחר לגמרי

עד אז עבדתי בעיקר עם מיקרו-מחשבים, שבהם הכלל הראשון היה פשוט:

לפני שנוגעים בציווד — מנתקים מהחשמל

ציווד לייצור מוליכים למחצה התגלה כעולם שונה לחלוטין.

1.12 Fab-8 ירושלים — הקמה, ספקנות וביקורת – איסי גייר

בשלב התכנון וההקמה של Fab-8, הפרויקט עמד בפני ספקנות וביקורת ציבורית ניכרת בישראל. **מוקד הביקורת:** חלק גדול מהביקורת התמקד במענקים הממשלתיים והתמריצים הכספיים שניתנו לאינטל. הדיון על מענקים אלה פרץ עוד בשלב התכנון של Fab-8, תוך כדי המשא ומתן בין אינטל לממשלת ישראל. הטענה היתה שהיה ניתן בכסף הזה להקים הרבה עסקים קטנים. היו שטענו כי מפעל אינטל לא ייצר מוצרי מוליכים למחצה אמיתיים. אחת השיחות הבלתי-נשכחות ביותר שקימתי עם חבר ילדות שלי מבית ספר יסודי שעבד במשרד מבקר המדינה של ישראל — הוא טען כי מפעל אינטל ישמש בעיקר כ"מקלט מס" לחברה אמריקאית, וכי מה שיקרה בפועל הוא שמוצרים פשוט ישלחו הלוך ושוב כדי ליצור מראית עין של ייצור, שימש כמקלט מס לחברה אמריקאית.

"לא שכחתי לעולם את העלבון שהיה מרומז בדברים ההם." — איסי גייר, Intel Fab-8 ירושלים

דב פרוהמן בחר שלא להגיב פומבית לביקורת מסוג זה. במקום זאת, אמר:

"הם יגידו מה שיגידו, ואנחנו נעשה מה שצריך לעשות." — דב פרוהמן

ההיסטוריה הוכיחה שהוא צדק.

1.13 Fab-8 — פרטים טכניים ותפעוליים

Fab-8 הוקמה בשכונת הר חוצבים בירושלים — שכונת התעשייה שהפכה לאחר מכן ל"עמק הסיליקון של ירושלים". מפעל ייצור שבבי הסיליקון החל לפעול בסוף שנות ה-80, ויצרה שבבי EPROM ומוצרי זיכרון לשוק העולמי.

הערות	פרטים	פרמטר
מפעל ייצור שבבי סיליקון ראשון של אינטל מחוץ לארה"ב	הר חוצבים, ירושלים Fab 8	מיקום
לאחר שנות בנייה והכשרה	אמצע שנות ה-80	התחלת פעילות
כולל תהליכי ion implantation	EPROM, מוצרי זיכרון	טכנולוגיה ראשונה
גם סמנכ"ל אינטל הגלובלית	דב פרוהמן (עד 2001)	מנהל כללי
עדכון לאורך שנות ה-90	"4 → "6 (100→150mm)	גודל וויפר
שילוב ייחודי לא נפוץ בפאב	ייצור + R&D	Fab-8 vs Fab ראשית
הרחבה לדרום ישראל	קריית גת, 1995	פרויקט Fab-11

1.14 עדות אישית: תפקידי ב-Fab-8, 1984–1995

הכותב שירת כמהנדס ציוד ותהליך בתחום אימפלנטציית יונים ב-Intel Fab-8 ירושלים, בשנים 1984–1995. לאחר הכשרה ב-Fab-5 אורגון וב-Fab-7 אלברקי ניו מקסיקו חזרתי לירושלים והתקנתי את המכונה הראשונה במיפעל, משתיל יונים Nova-10-8 להשתלות כבדות כחלק מהצוות הראשוני שהקים את מפעל Fab 8.

אחד האתגרים המרכזיים בתחום האימפלנטציה בפאב-8 היה בדיוק הפיזיקה שמאמר זה עוסק בה: **טעינת מטען בשכבות הדיאלקטרי** — charging damage לשערי טרנסיסטורים כיצד קרן יונים יוצרת מטען על פרוסה.

- מתי המטען יוצר שדה מספיק חזק לגרום לפריצת Gate
- כיצד electron shower (מקלחת אלקטרונים) מנטרל ביעילות את הטעינה, אנחנו איפיינו והיטמענו את השימוש במיקלחת האלקטרונים למרות היתנגדות באינטל אורגון, בסופו שלדבר גם הם עשו זאת.

1.15 מלחמת המפרץ, קריית גת, ומורשת

ב-1991, בזמן מלחמת המפרץ הראשונה ומתקפות הטילים העיראקיים על ישראל, פרוהמן קיבל החלטה שהפכה לאגדה: **לשמור את אינטל ישראל פתוחה** — בניגוד להמלצות פיקוד העורף שביקש מכל עסק לא-חיוני לסגור. אינטל ישראל היתה אחד העסקים היחידים — וכמעט עסק הייצור היחיד — שנשאר פועל לאורך כל המלחמה. פרוהמן כתב על כך מאמר ב-*Harvard Business Review*.

ב-1995, פרוהמן הוביל את הקמת ה-**Fab-18 בקריית גת** — מפעל ייצור שבבי סיליקון שני בישראל, בדרום הארץ, בהשקעה של מיליארדי דולרים. כיום Intel Israel מייצרת מעבדים מתקדמים עם עשרות אלפי עובדים — תשתית ההיי-טק הישראלית שפרוהמן בנה לאורך שלושה עשורים.

1.16 פרסים והכרה — מורשת

שנה	פרס / הכרה	גוף מעניק
1982	IEEE Jack Morton Award	IEEE
1991	פרס ישראל למדעים מדויקים	מדינת ישראל
2008	IEEE Edison Medal	IEEE
2009	National Inventors Hall of Fame	ארצות הברית
2018	Computer History Museum Fellow	CHM

פרוהמן גם כתב ספר על מנהיגות: **Leadership the Hard Way** (Jossey-Bass, 2008), שבו הוא מסכם את עקרונות הניהול שלמד מחייו — מהשואה, דרך ההמצאה, ועד ניהול אינטל ישראל.

ההמשך הוא טכנולוגי למי שמעוניין להעמיק

המאמרים של דוב בסוף המאמר

פרק ב': מ-EPROM ל-Flash ול-SSD: ציר ההתפתחות המלא

הדרך מהמאמר של פרוהמן ולנצלינגר (1969) ועד לכונן SSD של 4 טרה-בייט שיושב במחשב הנייד שלכם אורכת כ-55 שנה של מחקר, הנדסה ומסחר. כל דור סיפר את אותה הסיפור בשפה שונה קצת: **מטען כלאו בשער מוליכים-למחצה = ביט מידע**. אבל בכל פעם — יותר צפוף, יותר מהיר, ויותר זול.

2.1 EPROM — ראשית הזיכרון הניתן לתכנות (1971)

ב-1971 פרסם דב פרוהמן את מאמרו על **FAMOS — Floating-Gate Avalanche-injection MOS**. הרעיון: במקום לסמוך על מלכודות אקראיות בניטריד (MNOS), **בונים שער צף** מפוליסייליקון מוליך, מבודד מכל צדדיו בתחמוצת SiO_2 . מטען מוזרק אל השער הצף על ידי **avalanche injection** — מנגנון שבו אלקטרונים מואצים בשדה גבוה ומוזרקים לתחמוצת. המטען נשמר שם ללא מקור חיצוני.

מחיקה ב-UV: כדי לאפס את הזיכרון, יש להאיר את השבב בקרינת אולטרה-סגול (~ 253.7 UV, ננומטר). הפוטונים בעלי אנרגיה של כ-4.9 eV מעניקים לאלקטרונים בשער הצף מספיק אנרגיה לצאת דרך התחמוצת. לכן ב-EPROM קלאסי היה

חלון קוורץ שקוף מעל הדיי — כדי שאור UV יוכל להגיע אל הרכיב. זמן מחיקה: ~ 20 דקות בנורת UV רגילה.

גרסה מסחרית: ב-1971 אינטל שיחררה את ה-

EPROM — **Intel 1702** של 2048 ביט (256 בייט). זה היה המוצר המסחרי הראשון מסוגו.

2.2 UV-EPROM — דור הפריחה (שנות ה-70 וה-80)

לאורך שנות ה-70 וה-80, ה-EPROM הפך לרכיב סטנדרטי בכל מערכת משובצת מחשב. כל אחד שעבד בתעשייה בתקופה זו — כולל בפאבריקות של אינטל — מכיר את חבילות ה-DIP הקרמיות עם החלון השקוף.

חידוש	קיבולת	רכיב	שנה
EPROM מסחרי ראשון	256 בייט	Intel 1702	1971
אחסון מוגבר	KB 1	Intel 2708	1975
5V יחיד	KB 2	Intel 2716	1977
צפיפות גבוהה	KB 8	Intel 2764	1979
CMOS תהליך	KB 64	Intel 27512	1984
טכנולוגיה בשלה	KB 128	27C010	1988

בעיית ה-UV: על אף הצלחת ה-EPROM, מחיקת UV הייתה מגבלה תפעולית קשה. צריך להוציא את הרכיב מהמעגל, להאיר אותו דקות רבות, ולהחזירו. לא ניתן לעדכן חלק מהתוכן — כל המחיקה היא מחיקת כל הזיכרון בבת אחת. ומה עם מכשירים שכבר יצאו לשוק? לא ניתן לעדכן.

2.3 EEPROM — מחיקה חשמלית (1978)

הפתרון הגיע ב-1978 עם פרסום ה-**EEPROM — Electrically Erasable PROM** (נכתב גם E²PROM). הרעיון: מחיקה חשמלית, ללא UV, על ידי הפעלת מתח גבוה בכיוון הפוך שגורם ל-Fowler–Nordheim tunneling של המטען הנוצח מהשער הצף.

החידוש המבני: ב-EEPROM מוסיפים תחמוצת

tunnel oxide דקה במיוחד (7–10 ננומטר) מתחת לשער הצף. שכבה זו דקה מספיק כדי לאפשר מנהור FN בשני הכיוונים — כתיבה ומחיקה — בשדות חשמליים מבוקרים. המכשיר אינו חייב לצאת מהמעגל.

רכיבים מרכזיים: EEPROM — Intel 2816 (1978) ראשון מסחרי. Intel 28F256 (1988). ה-EEPROM מצא שימוש רחב ב-BIOS של מחשבים אישיים, כרטיסי חכמים, ומערכות embedded. **חיסרון מרכזי:** כל תא דורש טרנזיסטור נוסף לבקרת המחיקה, מה שמגדיל את שטח הד"י ומייקר את ייצור.

2.4 NOR Flash — פריצת הדרך של Masuoka (1984)

ב-1984, מהנדס יפני בשם

Fujio Masuoka מטושיבה, הציג בכנס IEEE-IEDM ארכיטקטורה חדשה שהוא קרא לה

Flash Memory — מפני שהמחיקה מהירה כ"הבזק" (flash) של מצלמה. הרעיון המרכזי: **מוחקים בלוקים שלמים בבת אחת**, לא תא-תא. זה מפשט דרמטית את האלקטרוניקה ומאפשר צפיפות גבוהה בהרבה.

ארכיטקטורת NOR: בתאי NOR Flash, כל תא מחובר ישירות לקו הביט. זה מאפשר גישה אקראית מהירה לכל בייט — בדומה לזיכרון RAM לקריאה. זמן גישה: ~100 ננו-שנייה. NOR Flash שימש בעיקר לאחסון קוד — BIOS, firmware של טלפונים ניידים.

מאפיין	EEPROM	NOR Flash	הערה
מחיקה	תא-תא	בלוק (+64KB)	Flash מהיר יותר
גישה לקריאה	אקראית	אקראית מהירה	XiP אפשרי
צפיפות	נמוכה	בינונית	פחות טרנזיסטורים
שימוש עיקרי	הגדרות	Firmware/BIOS	קוד הפעלה

NAND Flash 2.5 — הצפיפות המנצחת (1987)

ב-1987 הציג Masuoka ארכיטקטורה שנייה:

NAND Flash. כאן התאים מחוברים בטור (כמו שערי NAND), לא במקביל. אין גישה אקראית ישירה לבייט — הקריאה מתבצעת בדפים (pages) של 4–16 KB. המחיקה מתבצעת בבלוקים של 64–256 עמודים.

היתרון המכריע: כל תא קטן בהרבה — פחות חיבורי מתכת, פחות שטח. NAND Flash אינו מתאים לאחסון קוד הרץ ישירות (XiP), אבל מושלם לאחסון נתונים בנפחים גדולים. זה הפך אותו ל"ים האחסון" של מוצרי הצרכנות.

SLC, MLC, TLC, QLC 2.6 — יותר ביטים לתא

עד שנות ה-90, כל תא Flash אגר **ביט אחד** — שתי רמות מתח (SLC — Single Level Cell). המהפכה הבאה: אחסון

מספר ביטים בתא אחד על ידי הבחנה בין רמות מתח שונות של השער הצף.

צפיפות	אמינות	רמות מתח	ביטים/תא	טכנולוגיה
נמוכה	גבוהה מאוד	2	1	SLC
בינונית	גבוהה	4	2	MLC
גבוהה	בינונית	8	3	TLC

QLC	4	16	נמוכה יחסית	גבוהה מאוד
PLC (ניסיוני)	5	32	נמוכה	מקסימלית

הפיזיקה: כדי לאחסן 4 ביטים בתא QLC, צריך להבחין בין 16 רמות מתח שונות של השער הצף. ההפרש בין רמה לרמה הוא עשרות מילי-וולט בלבד. זה דורש אלקטרוניקת שליטה מתוחכמת ואלגוריתמי קידוד מורכבים (ECC — Error Correction Code).

2.7 3D NAND — הפתרון לגבול הפיזיקלי (2007 ואילך)

בסביבות 2007–2010, ייצור ה-NAND Flash הגיע לגבול הפיזיקלי של ה-

scaling הדו-ממדי. ככל שהתאים מתקצרים מתחת ל-20 ננומטר —

הרעשים, ה-*interference* בין תאים סמוכים, ובעיות ה-*retention* הפכו בלתי נסבלות.

הפתרון:

לבנות לגובה. במקום לדחוס תאים ב-2D, עורמים שכבות של תאים אחת על גבי השנייה. טכנולוגיית

3D NAND (או V-NAND — Vertical NAND) מאפשרת עשרות עד מאות שכבות אנכיות. Samsung שחררה את 3D V-NAND הראשון ב-2013. כיום, Samsung, Micron, SK Hynix, Kioxia ואחרים מייצרים 3D NAND עם 176–238 שכבות ויותר.

שנה	טכנולוגיה	שכבות	צפיפות אופיינית
2013	Samsung V-NAND ראשון	24	Gb/die 128
2016	3D TLC שכבה 64	64	Gb/die 256
2019	128–96 שכבות	128–96	Gb/die 512
2021	+176 שכבות	176	Tb/die 1
2023	+238 שכבות	238	Tb/die 2
2025 (צפוי)	+300 שכבות	+300	Tb/die 4

2.8 כונני SSD — מ-Flash לאחסון מסחרי

ה-SSD (Solid State Drive) הוא שילוב של שבבי NAND Flash, בקר (controller) חכם, ו-DRAM cache, שמתחזה לכונן דיסק רגיל מבחינת הממשק. ה-SSD הראשונים הופיעו בשנות ה-

90 (SunDisk, M-Systems — חברה ישראלית!), אך פרצו לשוק הצרכנים רק עם SSD של Intel ו-Samsung ב-2008–2009.

המבנה הפנימי של SSD:

- שבבי NAND Flash — הזיכרון עצמו (SLC/MLC/TLC/QLC)
- Controller — מעבד ייעודי שמנהל mapping, wear leveling, garbage collection, ECC
- DRAM — מטמון (cache) של כתובות וקריאות/כתיבות זמניות
- ממשק — SATA, PCIe/NVMe, M.2 — הקשר אל שאר המערכת

מאפיין	HDD מסורתי	SSD SATA	SSD NVMe
זמן גישה	10–5 מ"ש	0.1 מ"ש	0.05–0.02 מ"ש
מהירות קריאה	100–200 MB/s	500–600 MB/s	3,000–7,000 MB/s
חלקים נעים	כן (ראש+דיסק)	אין	אין
צריכת חשמל	5–10W	2–4W	3–6W
עמידות זעזועים	נמוכה	גבוהה מאוד	גבוהה מאוד

2.9 Wear Leveling — ניהול הבלאי

אחד האתגרים הגדולים של Flash: כל תא יכול לעמוד בכמות מוגבלת של מחזורי כתיבה-מחיקה:

- SLC Flash מחזורים: ~100,000
- MLC Flash מחזורים: ~10,000
- TLC Flash מחזורים: ~3,000
- QLC Flash מחזורים: ~1,000

בלאי זה נובע בדיוק מהפיזיקה של MNOS ו-Floating Gate: כל מחזור כתיבה-מחיקה גורם להזרקה מטען נוספת דרך ה-tunnel oxide, שיוצרת מלכודות חדשות, מחלישה את שכבת SiO_2 , ומקטינה בהדרגה את זמן תוחלת הזיכרון.

הפתרון הוא **Wear Leveling** — אלגוריתם שמבטיח שהכתיבות מתפזרות באופן שווה על כלל התאים, לא מתרכזות באזורים פעילים. זו אחת הפונקציות המרכזיות של ה-controller ב-SSD.

2.10 ECC — תיקון שגיאות כתנאי הישרדות

ב-QLC Flash, ההבדל בין 16 רמות מתח הוא כה קטן, שרעשים תרמיים ו-interference יכולים לגרום לשגיאות קריאה. לכן **ECC (Error Correction Code)** אינו אופציה — הוא הכרחי לפעולה תקינה.

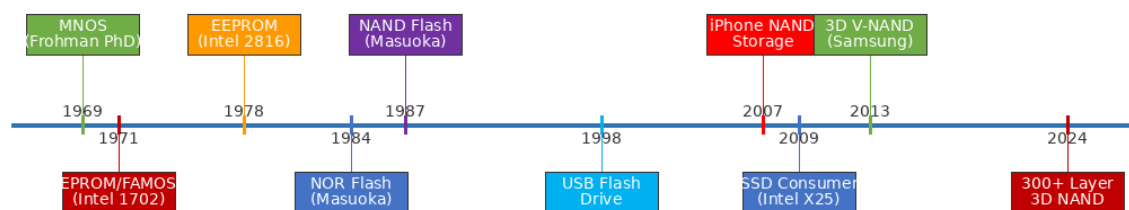
ב-SSD מודרניים נעשה שימוש באלגוריתמי ECC מתקדמים כגון

שיתוף פעולה הדוק בין פיזיקת ה-Flash לבין הנדסת מידע. **LDPC (Low-Density Parity-Check)**, שמוגלים לתקן מאות ביטים שגויים בכל דף. זהו

2.11 ציר זמן מלא: מ-MNOS ל-SSD

שנה	טכנולוגיה / אירוע	משמעות
1969	MNOS — מאמר פרוהמן ולנצלינגר	בסיס פיזיקלי לזיכרון לא-נדיף
1971	FAMOS / EPROM — אינטל	Floating Gate ראשון מסחרי
1978	EEPROM — אינטל 2816	מחיקה חשמלית ראשונה
1984	NOR Flash — Masuoka/Toshiba	מחיקת בלוקים, Flash ראשון
1987	NAND Flash — Masuoka/Toshiba	ארכיטקטורה לצפיפות גבוהה
1991	SanDisk FlashDisk 20MB	Flash drive מסחרי ראשון
1995	M-Systems DiskOnChip (ישראל)	Flash כדיסק לוגי
1998	USB Flash Drive (M-Systems)	אחסון נייד מהפכני
2007	NAND Flash storage — iPhone כ-	Flash בסמארטפונים
09–2008	SSD מסחרי — Intel X25-M	SSD לשוק הצרכנים
2013	Samsung 3D V-NAND	מעבר לממד השלישי
2019	QLC 3D NAND בנפח +4TB	כונני SSD צרכניים ענקיים
+2023	238 שכבות 3D NAND	גבול הטכנולוגיה הנוכחי

Non-Volatile Memory Evolution 1969 → 2024



איור 3 — ציר הזמן של התפתחות זיכרון לא-נדיף: מ-1969 MNOS עד 2024 3D NAND

2.12 הרלוונטיות של פיזיקת MNOS לכל אלה

מה שמדהים הוא שהפיזיקה הבסיסית של כל הטכנולוגיות הללו — EPROM, EEPROM, NOR — Flash, NAND Flash, 3D NAND — אינה שונה ביסודה מהמאמר של 1969:

- tunnel oxide דרך Fowler–Nordheim tunneling — בכתיבה ובמחיקה
- Floating Gate — אחסון מטען על שער מוליך מבודד
- שינוי מתח הסף כביטוי לאחסון המידע
- oxide ב-Trapping כגורם הבלאי העיקרי (נוצר בכל מחזור כתיבה)
- Retention — תלות בעומק המלכודות ובגיל ה-oxide

כאשר מהנדס Flash מדבר על *oxide degradation* או *charge loss* או *threshold voltage distribution* — הוא מדבר בדיוק בשפה שמאמר פרוהמן ולנצלינגר ניסח לראשונה.

ה-SSD שמחובר עכשיו למחשב שלכם מכיל מיליארדי תאי NAND Flash, שכל אחד מהם פועל לפי אותה פיזיקה בסיסית שתיאר מאמר בן 12 עמודים שפורסם בשנה שבה נחת האדם על הירח.

פרק ג': הבסיס הפיזיקלי — מבנה MNOS, מנגנוני הולכה וזיכרון לא-נדיף

The Physics: MNOS Structure, Charge Transport & Non-Volatile Memory

1. פרטים ביבליוגרפיים

פרט	מאמר	דוקטורט
קותרת	Charge Transport and Storage in MNOS Structures	and Its Memory... Applications
מחברים	& Frohman-Bentchkowsky Lenzlinger	Dov Frohman-Bentchkowsky
מוסד / כתב עת	Journal of Applied Physics	UC Berkeley

שנה	1969	1969
כרך / גיליון	Vol. 40, No. 8	—
עמודים	3319–3307	—

2. הקשר ההיסטורי

בסוף שנות ה-60 של המאה ה-20, תעשיית המוליכים למחצה עמדה בפני אחד האתגרים הגדולים שלה: כיצד לשמר מידע בתוך רכיב MOS גם לאחר ניתוק המתח. הזיכרונות הקיימים אז — *core memory* מגנטיים, מסכי פריט — היו גדולים, יקרים ואיטיים. הצורך ב-**non-volatile memory** ישירה על הרכיב היה ברור.

דב פרוהמן, שהשלים את דוקטורט בברקלי ב-1969 תחת הנחיית Martin H. Graham, בחר לחקור מבנה מיוחד שבו מוסיפים שכבת ניטריד סיליקון (Si_3N_4) מעל תחמוצת הסיליקון הרגילה. הניטריד, כפי שיתברר, אינו מבודד אידיאלי — ויש לכך משמעויות מרחיקות לכת.

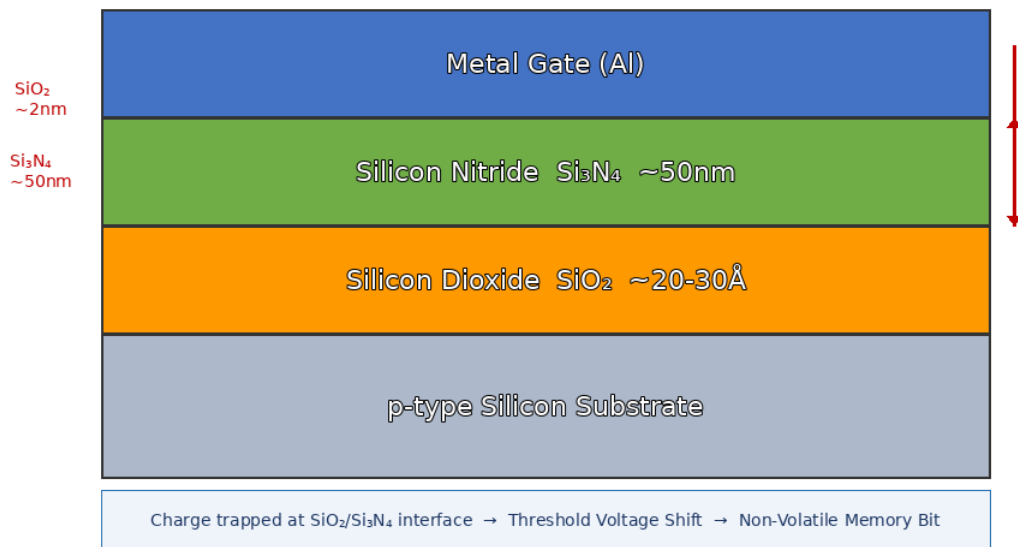
3. מבנה MNOS: ארכיטקטורה ותכונות

3.1 הארכיטקטורה הכללית

מבנה MNOS הוא שכבות דיאלקטריות כפולות מעל מצע הסיליקון:

תפקיד	חומר	שכבה
מגע חיצוני, הפעלת שדה אחסון מטען (trapping)	מתכת (Al) — Si_3N_4 ניטריד סיליקון	שער דיאלקטרי עליון
מנהור ובקרת זרם	SiO_2 — תחמוצת סיליקון דקה	דיאלקטרי תחתון
ספק נשאי מטען	סיליקון (n-type או p-type)	מצע

MNOS Layer Structure



איור 1 — מבנה שכבות MNOS ומנגנון לכידת המטען בממשק SiO_2/Si_3N_4

ההבדל מ-MOS רגיל (Metal/Oxide/Silicon) הוא תוספת שכבת ניטרید. תוספת זו אינה קטנה: היא הופכת את כל המבנה ממבנה פסיבי (מבודד) לפעיל — מסוגל לאגור מידע.

3.2 שכבת SiO₂ — "שער המנהור"

שכבת התחמוצת ב-MNOS היא דקה במיוחד — בסדר גודל של 20–30 ångström בניסויים המוקדמים. הדקות הזאת אינה מקרית:

- כאשר השדה החשמלי גבוה, התחמוצת מאפשרת מעבר מטען דרך מנהור קוונטי.
- כאשר השדה נמוך (בזמן שמירה), התחמוצת מהווה מחסום יעיל המונע בריחת מטען.
- כמו "ברז" שנפתח בכוח גבוה ונסגר בכוח נמוך — זהו בסיס מחזור הכתיבה/שמירה.

אם התחמוצת עבה מדי — אי אפשר לכתוב. אם דקה מדי — המטען בורח מהר. המינון המדויק של עובי התחמוצת הוא מהגורמים הקריטיים בתכנון MNOS.

3.3 שכבת Si₃N₄ — ה"מחסן"

ניטרید הסיליקון אינו מבודד מושלם. הוא מכיל צפיפות גבוהה של מצבי מלכודת (traps) — פגמים מבניים, קשרים תלויים (dangling bonds), ומצבי ביניים. בטכנולוגיית MOS קלאסית, traps נחשבים לבעיה שיש למזער. ב-MNOS, המלכודות הן המנגנון עצמו — הן מהוות את אמצעי האחסון.

האלקטרונים הנכנסים לניטרید נלכדים במלכודות הללו, ושם הם נשארים גם ללא מתח חיצוני. זוהי הגדרת הזיכרון הלא-נדיף: מידע נשמר ללא מקור אנרגיה.

4. מנגנוני הולכת מטען

המאמר מנתח בפירוט את הדרכים שבהן מטען עובר דרך המבנה הדו-שכבתי. ההבנה של מנגנונים אלה חיונית הן לתכנון זיכרון יעיל והן לניבוי כשלים ו-degradation.

4.1 Fowler–Nordheim Tunneling בתחמוצת

בשדה חשמלי גבוה, התחמוצת מאפשרת מנהור מסוג **Fowler–Nordheim**. בניגוד להולכה אוהמית רגילה, כאן הגורם המרכזי הוא לא מוליכות החומר אלא **הסתברות קוונטית** לחצות את המחסום.

מה קורה פיזיקלית? ללא שדה, מחסום האנרגיה של אלקטרון בחומת SiO_2 הוא גבוה ועבה — כמעט אין מעבר. כאשר מפעילים שדה חשמלי חזק, המחסום נעשה **משולש** ודק יותר בצד אחד. האלקטרון, בהיותו גל קוונטי, יכול **למנהר** דרך החלק הדק. הסתברות המנהור עולה דרמטית עם עלייה בשדה.

משמעות מעשית: יש **סף מעשי** — מתחת לשדה מסוים כמעט אין כתיבה; מעליו, הכתיבה מתחילה בחדות. זה מאפשר פעולה דיגיטלית: מצב שמירה (שדה נמוך) מול מצב כתיבה (שדה גבוה).

4.2 Poole–Frenkel Conduction בניטריד

בניטריד, המנגנון הדומיננטי הוא **Poole–Frenkel** (המכונה גם Frenkel–Poole). זהו תהליך תרמי-שדה: האלקטרון נמצא בתוך בור פוטנציאל (מלכודת), ושדה חשמלי חיצוני מוריד את מחסום היציאה מהמלכודת.

הבדל מ-Fowler–Nordheim: ב-FN האלקטרון ממנהר דרך מחסום שלם. ב-Poole–Frenkel האלקטרון קיים כבר בתוך הניטריד (כלוא במלכודת), והשדה מעניק לו אנרגיה לצאת ממנה. לכן F -P רגיש יותר לטמפרטורה ולמיקום המלכודות.

השלכה לזיכרון: אותן מלכודות שמאפשרות את הלכידה גם מאפשרות הולכה בשדה גבוה. כלומר — בכתיבה, מטען זורם דרך המלכודות. בשמירה, המלכודות "מחזיקות" אותו. זוהי הדואליות הבסיסית של הניטריד.

4.3 Hopping Conduction

בשדות נמוכים יחסית ובטמפרטורות מתונות, יכול להתקיים **hopping conduction**: אלקטרון "קופץ" ממלכודת סמוכה למלכודת, ללא עליה לפס ההולכה. תהליך זה איטי, אך חשוב בהבנת הזליגה ה

drift לאורך זמן — כלומר האובדן ההדרגתי של מידע ב-MNOS.

4.4 סיכום מנגנוני הולכה בטבלה

תנאים	מיקום	מנגנון
שדה גבוה מאוד (כתיבה)	SiO ₂	Fowler–Nordheim tunneling
שדה בינוני-גבוה, T > 0	Si ₃ N ₄	Poole–Frenkel
שדה נמוך, T מתון	Si ₃ N ₄	Hopping
שדות גבוהים מאוד (משני)	Si ₃ N ₄	Field emission (מנהור ממלכודות)

5. לכידת מטען בממשק SiO₂/Si₃N₄ — לב המאמר

זהו החידוש המרכזי של המאמר. הפיזיקה אומרת שמטען אינו מתחלק באופן אחיד בכל הניטריויד. הוא מצטבר, באופן מועדף, בממשק שבין SiO₂ לבין Si₃N₄. למה? כי שני החומרים מוליכים אחרת.

5.1 הטיעון הפיזיקלי

עבור זרם רציף לזרום דרך שתי שכבות, חוק רציפות הזרם ($J = const$) מחייב שהשדה בכל שכבה "יתאים" לזרם הנדרש. מאחר שמנגנוני הולכה שונים ממשיכים שדות שונים, הפתרון של המערכת הוא הצטברות מטען בממשק — כדי לפרק את הדרישות הנפרדות של כל שכבה לשדה מקומי מתאים.

במילים אחרות: המטען בממשק הוא לא תוצאת לוואי. הוא תוצאה פיזיקלית בלתי נמנעת של הבדל המנגנונים בין שתי שכבות.

5.2 ה-Sheet Charge ושינוי מתח הסף

את המטען הכלוא ניתן לתאר כ

sheet charge בממשק — צפיפות מטען משטחית [C/cm²]. מטען זה יוצר שינוי בפוטנציאל:

$$\Delta V_{fb} = -Q / C_{ox}$$

כאשר C_{ox} הוא הקיבוליות של שכבת התחמוצת. השינוי ΔV_{fb} מתבטא כהזזה של עקומת C–V — זהו הביטוי הנמדד של הזיכרון. חשוב לציין: ככל שהמטען קרוב יותר לסיליקון, השפעתו חזקה יותר. מטען בממשק SiO₂/Si₃N₄ יעיל יותר מאשר מטען עמוק בניטריויד.

6. מחזור הזיכרון: כתיבה, שמירה, קריאה, מחיקה

6.1 כתיבה (Program)

הפעלת מתח שער גבוה (לדוגמה, $20V+$ עד $40V+$) יוצרת שדה חשמלי חזק בתחמוצת. השדה מאפשר **Fowler-Nordheim tunneling** מהסיליקון לתחמוצת. המטען עובר לניטריד ונלכד בממשק. התהליך תלוי ב:

- גובה המתח
- משך הפולס
- עובי התחמוצת
- צפיפות המלכודות
- טמפרטורה

6.2 שמירה (Retention)

ללא מתח חיצוני, השדה בתחמוצת נמוך מאוד. המחסום מחזיר את עצמו לצורה עבה. המטען נשאר כלא במלכודות הניטריד. Retention טיפוסי של MNOS — שנים (בתנאים נורמליים). האויבים של Retention הם:

- טמפרטורה גבוהה — מאיצה שחרור תרמי ממלכודות
- שדה שיורי — יכול לגרום לזליגה הדרגתית
- מלכודות רדודות — נשמרות זמן קצר יותר

6.3 קריאה (Read)

הקריאה מתבצעת על ידי מדידת מתח הסף, בדרך כלל ללא הרסנות — מפעילים מתח שער נמוך ומודדים אם הטרנזיסטור מוליך או כבוי. המתח הנמוך אינו מספיק לשנות את המטען הכלוא, לכן הקריאה אינה פוגעת בנתון.

6.4 מחיקה (Erase)

מחיקה ב-MNOS מתבצעת חשמלית — על ידי הפעלת מתח הפוך. השדה הפוך מוציא מטען (או מזריק מטען הפוך) ומחזיר את מתח הסף לקרוב למצב המקורי.

זהו יתרון משמעותי של MNOS לעומת EPROM: ה-EPROM המקורי של פרוהמן (Floating Gate) דרש מחיקה בקרינת UV (אור אולטרה-סגול), שדרשה פתח בחבילה. MNOS לא דרש זאת.

7. שיטות מדידה: C-V ו-I-V

7.1 מדידת C-V

מדידת קיבול-מתח (C-V) היא הכלי המרכזי לניתוח MNOS. עקומת C-V מראה את הקיבול בין השער למצע כפונקציה של מתח השער.

(ב-MOS רגיל)

ללא *(trapping)*, הסריקה קדימה ואחורה נותנות עקומות זהות. ב-MNOS, הסריקה קדימה מזריקה מטען שנלכד, ולכן הסריקה חזרה נעה על מסלול אחר. נוצר **hysteresis loop**. רוחב ה-loop הוא מדד לחלון הזיכרון.

C-V shift: כאשר מפעילים פולס כתיבה לפני המדידה, העקומה כולה זזה — ימינה (מטען שלילי) או שמאלה (מטען חיובי). ΔV הנמדד פרופורציוני למטען הכלוא.

7.2 מדידת I-V

מדידת זרם-מתח (**I-V**) משמשת לניתוח מנגנוני הולכה. ציור ה

$\ln(J)$ מול $E\sqrt{}$ (שורש השדה) מאפשר לזהות **Poole-Frenkel**: אם התלות לינארית — זה אות ל-PF conduction. ציור $\ln(J/E^2)$ מול $E/1$ מאפשר לזהות **Fowler-Nordheim** (תלות לינארית על גרף זה — אופייני ל-FN).

8. בעיות ומגבלות של MNOS

כמו כל טכנולוגיה, MNOS מלווה בבעיות מהותיות שהגבילו את שימושו המסחרי:

8.1 פיזור מלכודות

המלכודות בניטריד אינן אחידות — לא בעומק ולא בצפיפות. לכן התגובה לפולס כתיבה אינה זהה מתא לתא. זה מקשה על שליטה מדויקת ב"חלון הזיכרון".

8.2 Drift — שינוי הדרגתי

לאחר כתיבה, מתח הסף המדוד אינו יציב לחלוטין. חלק מהמטען יכול לזוז לאט — *drift* — ולגרום לשינוי הדרגתי בחלון הזיכרון. זה מצריך **refresh** תקופתי בחלק מהיישומים.

8.3 Endurance — בלאי

בכל מחזור כתיבה/מחיקה מופעלים שדות חשמליים חזקים בדיאלקטרים. לאורך זמן נוצרות מלכודות חדשות, חלק מהמלכודות הקיימות משתנות, ויכול להיווצר **charge build-up** קבוע שאינו נמחק. תהליך זה מגביל את *endurance* — מספר מחזורי הכתיבה.

8.4 תלות בטמפרטורה

ה-Poole-Frenkel ו-hopping תלויים בטמפרטורה. לכן מאפייני ה-MNOS (מהירות כתיבה, retention, חלון זיכרון) משתנים עם הטמפרטורה. **זו בעיה מרכזית** ביישומים תעשייתיים שדורשים טווח טמפרטורות רחב.

9. MNOS מול Floating Gate EPROM: השוואה

המעבר של פרוהמן מ-MNOS ל-Floating Gate EPROM לא היה נטישה של הרעיון — אלא שיפור הנדסי על אותה תובנה בסיסית.

מאפיין	MNOS	EPROM (Floating Gate)
אחסון מטען	מלכודות בניטריד (אקראי)	שער צף מוליך (מוגדר)
מחיקה	חשמלית (מתח הפוך)	UV (דור ראשון)
יציבות חלון	בינונית (drift)	גבוהה יותר
שליטה בכתיבה	מורכבת	פשוטה יותר
ייצור	פשוט יחסית	מורכב יותר (פוליסייליקון)
מחיקה במיקרו	אפשרית	UV — חייב חלון

הקפיצה הרעיונית המשותפת: בשני המקרים, מטען כלוא בתוך מבנה שער של MOS משנה את מתח הסף. ההבדל הוא רק היכן שמים את המטען — במלכודות דיאלקטרי (MNOS) או על מוליך מבודד (Floating Gate).

10. ציר זמן היסטורי

שנה	אירוע	משמעות
1967–69	מחקר MNOS בברקלי ו-Fairchild	בסיס ניסויי ותיאורטי
יולי 1969	מאמר JAP על MNOS	פרסום מדעי מרכזי
1969	דוקטורט של פרוהמן בברקלי	"...and Its Memory... Applications"
1970	מאמר הרחבה ב-Proc. IEEE	MNOS transistor — יישומים
1971	FAMOS ו-EPROM באינטל	Floating Gate — מסחור
1980+	EEPROM, Flash	מחיקה חשמלית מלאה
היום	NAND Flash, 3D NAND	ירוסה ישירה

11. קשר לעבודה בתעשייה: חיבור לאימפלנטציה

פיזיקת MNOS אינה נשארת בגדר מחקר אקדמי — היא נוגעת ישירות לתהליכי ייצור שבהם נוצרת טעינה לא רצויה של מבודדים. מהנדסי תהליך שעבדו עם משתילי יונים ותהליכי פלזמה מכירים את התופעה: *charging*.

11.1 Gate Oxide-b Damage-ו Charging

במהלך השתלת יונים, מיוחדת כשיש שכבות מבודדות על הוויפר (כגון שערי טרנזיסטורים MOS), יכול להיווצר פוטנציאל שטח גבוה. הסיבה: צבירת מטען מהקרן על מבודדים. פוטנציאל גבוה זה יוצר שדה חשמלי בתחמוצת השער — בדיוק כמו בזמן כתיבה ב-MNOS. אם השדה גבוה מספיק, יכול להיות:

- מנהור Fowler–Nordheim בתחמוצת השער
- יצירת מלכודות חדשות ב-oxide
- Threshold shift של הטרנזיסטור
- breakdown מקומי

11.2 Electron Shower — הפתרון

הפתרון הקלאסי בתעשייה הוא **electron shower** ("מקלחת אלקטרונים") — פליטת אלקטרונים תרמיים מפילמנט לעבר הוויפר, לנטרול המטען החיובי שנצבר. הבנת הצורך בכך — ומידת האפקטיביות שלו — מחייבת הבנת פיזיקת MNOS: כמה מטען צבור, איפה, ומה השפעתו.

כלומר: מהנדס אימפלנטציה שמבין MNOS מבין גם מדוע טעינה בזמן תהליך היא בעיה, ומדוע ניטרול מטען הוא הכרחי — לא רק נוהל שגרותי.

12. סיכום

המאמר של Frohman-Bentchkowsky ו-Lenzlinger (1969) הוא יצירה מדעית יסודית שמחברת בין:

- פיזיקת קוונטים (מנהור Fowler–Nordheim)
 - פיזיקת מצב מוצק (Poole–Frenkel, trapping)
 - הנדסת התקנים (MOS, threshold voltage)
 - רעיון חדשני (מטען כלוא = זיכרון)
- מדובר ב-24 עמודים מדעיים שנכתבו על ידי חוקר בן 28 בתחילת קריירה. פרוהמן לקח תופעה שנחשבה לבעיה — trapping בדיאלקטרי — הפך אותה לרכיב. זו החשיבה שמביאה לפריצות דרך. שנתיים לאחר מכן, ב-1971, אותה חשיבה הולידה את ה-EPROM — רכיב שיכנס לכל מחשב אישי, לכל מחשב שלוחה, ולכל פלאפון — מיליארדי רכיבים הנושאים את ה-DNA הפיזיקלי של הדוקטורט מברקלי 1969.

מאמר 1969

AIP | Journal of
Applied Physics

Charge Transport and Storage in MetalNitrideOxideSilicon (MNOS) Structures

D. FrohmanBentchkowsky and M. Lenzlinger

Citation: *J. Appl. Phys.* **40**, 3307 (1969); doi: 10.1063/1.1658181

View online: <http://dx.doi.org/10.1063/1.1658181>

View Table of Contents: <http://jap.aip.org/resource/1/JAPIAU/v40/i8>

Published by the American Institute of Physics.

Additional information on J. Appl. Phys.

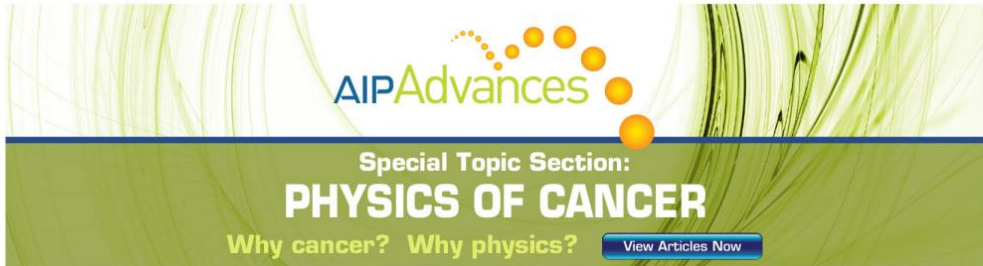
Journal Homepage: <http://jap.aip.org/>

Journal Information: http://jap.aip.org/about/about_the_journal

Top downloads: http://jap.aip.org/features/most_downloaded

Information for Authors: <http://jap.aip.org/authors>

ADVERTISEMENT



AIPAdvances

Special Topic Section:
PHYSICS OF CANCER

Why cancer? Why physics? [View Articles Now](#)

Charge Transport and Storage in Metal-Nitride-Oxide-Silicon (MNOS) Structures*

D. FROHMAN-BENTCHKOWSKY AND M. LENZLINGER

Fairchild Semiconductor, Research and Development Laboratory, 4001 Miranda Avenue, Palo Alto, California 94304

(Received 19 February 1969; in final form 31 March 1969)

A simple physical model that predicts charge accumulation at the dielectric interface of metal-nitride-oxide-silicon (MNOS) structures is proposed and verified experimentally. The model is based on the presence of steady-state current flow in the dielectric structure. Interface-charge accumulation is shown to be determined by the requirement for continuity of current through the structure under steady-state conditions. Continuity of current is established by accumulation of either positive or negative charge for a given polarity of charging voltage, depending on the relative current-field characteristics of the silicon nitride and silicon dioxide layers. Due to the exponential nature of the current-field characteristics, the time required to reach steady state is a strong function of the applied charging voltage. This leads to the observed charge storage property of MNOS devices. The hysteresis characteristic observed in MNOS structures is shown to be time-dependent with a tendency to merge into a single-valued dependence of accumulated charge on charging voltage as the steady-state condition is approached. The validity of the theoretical model for both steady-state and transient behavior is confirmed by current-voltage, capacitance-voltage, and turn-on measurements of MNOS capacitors and transistors for different dielectric thickness ratios and over a wide temperature range. The underlying concept that charge accumulation establishes current continuity in a two-layer dielectric structure should be valid, in general, for any two-dielectric structure.

I. INTRODUCTION

Metal-nitride-oxide-silicon (MNOS) devices have been shown to exhibit charge storage characteristics.¹ As a result, considerable interest has been expressed in their potential application as memory elements.² The memory function of these devices has been attributed to trapping of charge at the nitride-oxide interface. Consequently, it becomes important to understand charge transport mechanisms in MNOS structures. This paper will present a consistent model of charge transport and storage in MNOS structures.

Previous investigations have shown that steady-state current flows in both metal-oxide-silicon (MOS)³ and metal-nitride-silicon (MNS)⁴ structures. It is proposed that steady-state current also flows through the MNOS structure and that the charge trapped at the nitride-oxide interface is determined by the balance of the currents in the two dielectrics. Since the current-field characteristics are different for the two materials, the fields will tend to adjust themselves by charge accumulation at the interface until current continuity is established. On application of a positive voltage to the gate electrode, this model predicts the possibility of accumulation of either positive charge or negative charge at the nitride-oxide interface, depending on the conduction characteristics and thicknesses of the silicon nitride and silicon dioxide layers. Although the MNOS structure is used as a reference, the model is applicable to the general case of any two-dielectric structure.

The validity of the theoretical model is confirmed by current-voltage, capacitance-voltage, and turn-on measurements of MNOS capacitors and transistors for different dielectric conduction characteristics and over a wide temperature range. Current measurements are consistent with results previously published for conduction in both the silicon dioxide,³ due to tunneling from the silicon into the oxide conduction band, and the silicon nitride,⁴ due to excitation from traps into the silicon nitride conduction band. Both the conventional mode of device operation (accumulation of negative charge at the nitride-oxide interface for a positive applied voltage) and the inverse mode of operation (accumulation of positive charge at the nitride-oxide interface for a positive applied voltage) have been observed. Since current flow is predicted to persist down to very low applied voltage levels, the hysteresis characteristic observed in MNS and MNOS structures should merge into a single line if true steady state is obtained, i.e., if the voltages are applied for a long period of time (hours and days). This tendency has been confirmed experimentally on all structures investigated.

II. CONDUCTION IN SILICON DIOXIDE AND SILICON NITRIDE

Consider an MNOS structure as shown in Fig. 1. The variation of the energy bands with distance in a direction normal to the surface, for positive and negative bias on the metal electrode with respect to the semiconductor substrate, is illustrated in Fig. 2. An investigation of charge transport and storage in this structure requires an understanding of conduction mechanisms in MOS and MNS structures. Results of studies on these structures have been reported.^{3,4}

The dominant current-transport mechanism in MOS structures was shown to be electrode-limited due to

* This paper was presented at the IEEE International Electron Devices Meeting, Washington, D.C., October 1968.

¹ H. C. Pao and M. O'Connell, *Appl. Phys. Lett.* **12**, 260 (1968).

² H. A. R. Wegener *et al.*, "The Variable Threshold Transistor, a New Electrically-Alterable, Nondestructive Read-Only Storage Device," IEEE Int. Electron Devices Meeting, Washington, D.C., October 1967.

³ M. Lenzlinger and E. H. Snow, *J. Appl. Phys.* **40**, 278 (1969).

⁴ S. M. Sze, *J. Appl. Phys.* **38**, 2951 (1967).

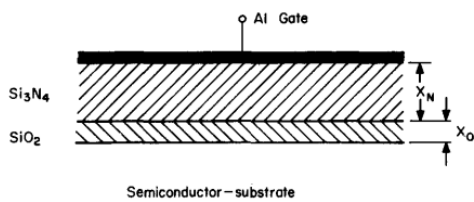


FIG. 1. Cross section of MNOS structure.

Fowler-Nordheim emission given by the following expression²:

$$J_0 = C_0 E_0^2 \frac{\pi c k T / E_0}{\sin(\pi c k T / E_0)} \exp\left(\frac{-E_1}{E_0}\right), \quad (1)$$

where E_0 is the electric field in the oxide, T is the temperature in °K, k is Boltzmann's constant, and E_1 , C_0 , c are characteristic constants depending on the barrier height of the interface. The temperature dependence of the Fowler-Nordheim current stems from the dependence on temperature of the number of electrons of a given energy incident on the barrier. The values of the constant terms in the current-density expression [Eq. (1)] for the oxide-silicon interface were found to be

$$C_0 = 10^{-5} \text{ A/V}^2, \\ E_1 = 2.54 \times 10^8 \text{ V/cm}, \\ c = 1.12 \times 10^{27} (\text{C} \cdot \text{cm})^{-1}.$$

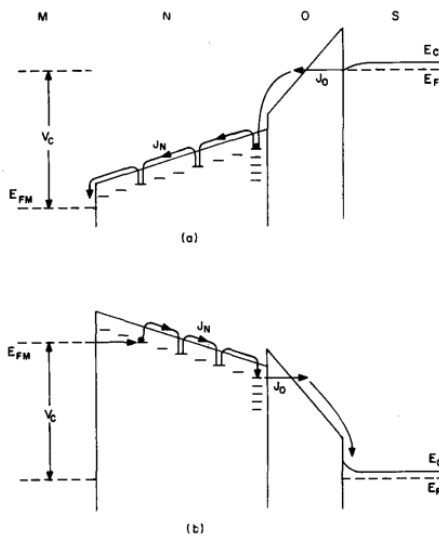


FIG. 2. Energy band diagram of MNOS structure under (a) positive bias and (b) negative bias.

When the thickness of the oxide is reduced to a value below which direct tunneling is possible from the semiconductor into the interface traps ($< 50 \text{ \AA}$), the expression for the current component in the oxide (J_0) has to be modified to account for the change in the conduction mechanism from Fowler-Nordheim emission to direct tunneling.⁵ This case has not been investigated in the present work. However, since the basic derivations of current transport in the MNOS structure are independent of the specific current mechanism in the oxide or the nitride, the model applies, in principle, to a structure which allows direct tunneling through the oxide.

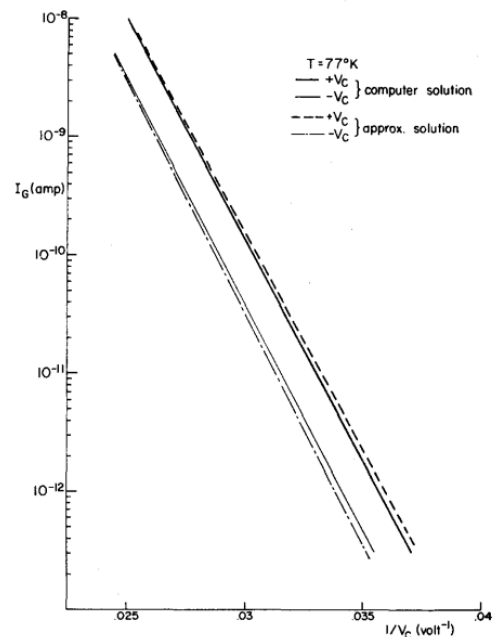


FIG. 3. Current as a function of charging voltage at low temperatures. $x_0 = 50 \text{ \AA}$, $x_n = 500 \text{ \AA}$, area = $1.3 \times 10^{-4} \text{ cm}^2$.

In MNS structures current transport was shown to be bulk-limited due to Schottky emission from traps above room temperature and field emission (tunneling) from the same traps into the silicon nitride conduction band at low temperatures. At low fields there is an additional ohmic component. The current-density expression consists of three terms:⁴

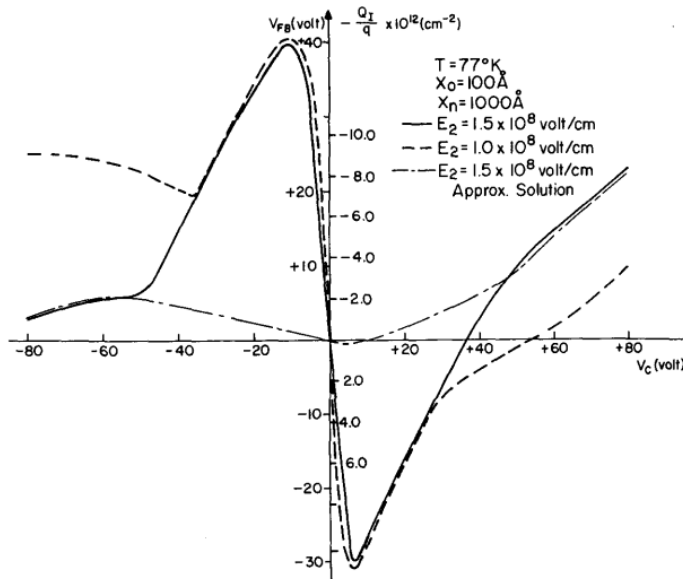
$$J_n = J_{n1} + J_{n2} + J_{n3}. \quad (2)$$

J_{n1} corresponds to field-enhanced thermal excitation (Poole-Frenkel effect) and is given by

$$J_{n1} = C_1 E_n \exp(-q\phi_1/kT) \exp[(q/kT)(\beta E_n)^{1/2}], \quad (3)$$

⁵ J. Shewchun, A. Waxman, and G. Warfield, *Solid-State Electron.* **10**, 1165 (1967).

FIG. 4. Equilibrium induced interface charge as a function of charging voltage at low temperature for two values of E_2 . The dot-dashed line represents the approximate solution for $E_2=1.5 \times 10^8$ V/cm.



where E_n is the electric field in the nitride, ϕ_1 is the depth of the trap-potential well, and C_1, β are characteristic constants depending on the trap level and the dielectric constant. J_{n2} is due to field emission of trapped electrons into the dielectric conduction band:

$$J_{n2} = C_2 E_n^2 \exp(-E_2/E_n) \quad (4)$$

E_2 and C_2 are characteristic constants depending on the trap level. J_{n3} results from hopping of thermally excited electrons between isolated trapping states.⁶

$$J_{n3} = C_3 E_n \exp(-q\phi_2/kT), \quad (5)$$

where $q\phi_2$ is the thermal activation energy.

The values of the constant terms in these current-density expressions for silicon nitride were found to be as follows:

$$\begin{aligned} C_1 &= 3.0 \times 10^{-9} \text{ A/V} \cdot \text{cm}, & \phi_1 &= 1.0 \pm 0.2 \text{ V}, \\ C_2 &= 3.5 \times 10^{-10} \text{ A/V}^2, & E_2 &= 1.2 \times 10^8 \text{ V/cm}, \\ C_3 &= 5.0 \times 10^{-14} \text{ A/V} \cdot \text{cm}, & \phi_2 &= 0.1 \text{ V}, \\ & & \beta &= 1.18 \times 10^{-7} \text{ V} \cdot \text{cm}. \end{aligned}$$

The deviation from the values reported in Ref. 4 is due to a difference in deposition conditions affecting the characteristic constants of the current-density expressions. In the derivations that follow, the values of the constants listed above (for the oxide and the nitride) will be assumed unless otherwise specified.

⁶ N. F. Mott and W. D. Twose, *Advan. Phys.* **10**, 107 (1961).

III. CONDUCTION AND STORAGE IN MNOS STRUCTURES

Based on the results in Sec. II, it is expected that current will flow in the MNOS structure (Fig. 1) on application of a bias voltage to the metal electrode. Figure 2(a) depicts the case of a positive metal electrode or gate voltage with respect to the semiconductor substrate. In this case the steady-state current consists of electrons tunneling from the semiconductor into the oxide conduction band, dropping into traps at the nitride-oxide interface, and being excited in turn into the nitride conduction band to the metal electrode. This corresponds to the current-transport mechanisms described in Eqs. (1) and (2), respectively. The description is limited to electron transport since evidence of hole transport in this structure has not been reported and no evidence for it has been found in this investigation. For the case of a negative metal bias [Fig. 2(b)], the steady-state current will similarly consist of electrons, which are excited from traps into the silicon nitride conduction band, tunneling through the nitride-oxide barrier into the oxide and semiconductor conduction band. Note that conduction in this case involves electron transport through the nitride-oxide interface barrier. It is assumed here that the transport mechanism is Fowler-Nordheim tunneling from the silicon nitride into the silicon dioxide conduction band. (The transport mechanism in this case could also be due to hole emission from the silicon into the silicon dioxide valence band. However, the two mechanisms could not be distinguished experimentally.) Consequently, Eq. (1) should apply with different values for C_0 and E_1 :

$$J_{0-} = C_0 E_0^2 \exp(-E_1/|E_0|) \quad (6)$$

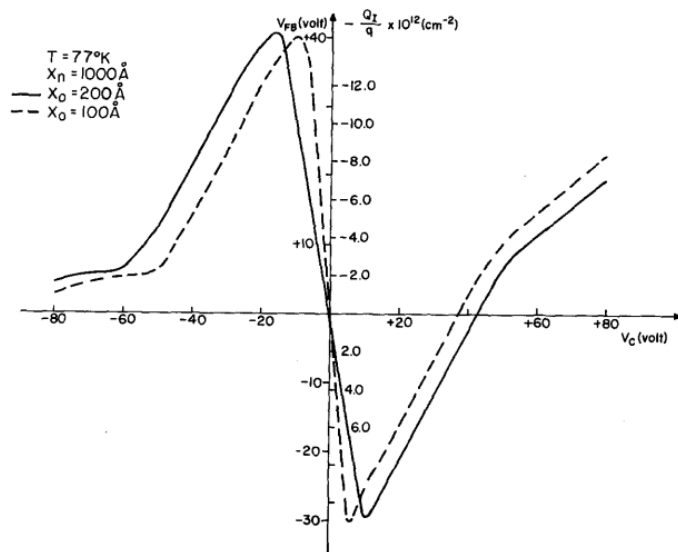


FIG. 5. Equilibrium induced interface charge as a function of charging voltage at low temperature for two values of oxide thickness.

Since the energy-barrier height of this interface has not been determined previously, the values of C_{0-} and E_{1-} were obtained from experimental measurements at low temperatures on MNOS structures, resulting in

$$C_{0-} = 9.0 \times 10^{-8} \text{ A/V}^2,$$

$$E_{1-} = 3.2 \times 10^8 \text{ V/cm.}$$

Because in this case the temperature dependence of the

number of carriers available for tunneling is not known, no temperature dependence was assumed for J_{0-} .

On application of a charging gate voltage (V_C), the fields in the oxide and the nitride (E_o, E_n) are determined initially by the dielectric constant ratio of the two dielectrics. The buildup of current through the structure will depend on the current-field characteristics of the nitride [Eq. (2)] and the oxide [Eq. (1)]. These characteristics, being different, result in a non-

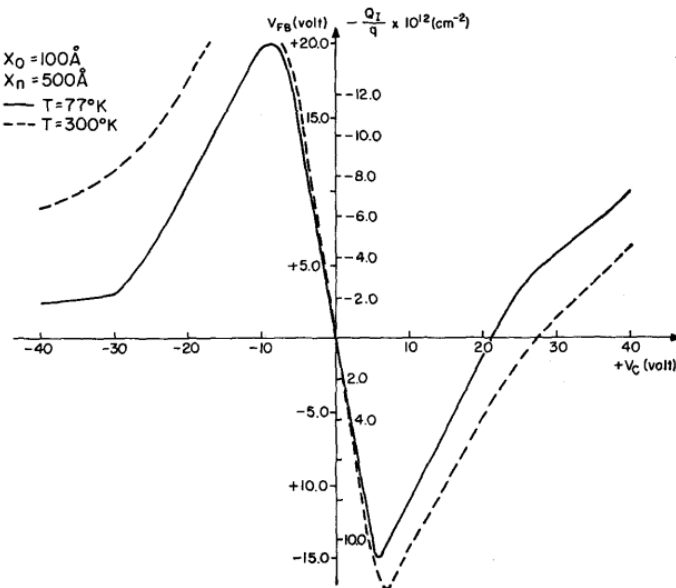


FIG. 6. Temperature dependence of equilibrium induced interface charge as a function of charging voltage.

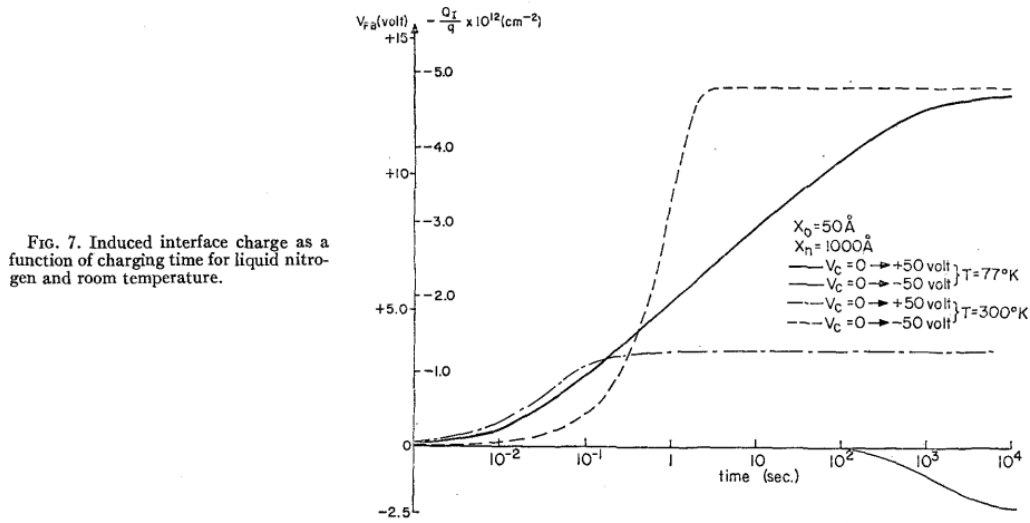


FIG. 7. Induced interface charge as a function of charging time for liquid nitrogen and room temperature.

equilibrium current discontinuity across the dielectric interface. The current discontinuity leads to charge accumulation at the interface, which in turn adjusts the electric field distribution until current continuity is established. This observation leads naturally to the steady-state and transient solutions for both the current and the charge. Note that the description of the interface trapping levels is only qualitative. The steady-state and transient solutions for charge transport in the MNOS structure do not require knowledge of the exact trap distribution.

A. Steady-State Analysis

The steady-state solution will consist of a determination of the electric fields in the two dielectrics (E_0, E_n) and the charge accumulated at the dielectric interface (Q_I). Current continuity in the steady state requires that for a given applied voltage

$$J_0 = J_n \tag{7}$$

Continuity of the electric flux leads to

$$K_0 E_0 - K_n E_n = Q_I / \epsilon_0 \tag{8}$$

where Q_I is the charge accumulation at the nitride-oxide interface, $K_0=3.9$ is the relative dielectric constant of silicon dioxide, K_n is the relative dielectric constant of silicon nitride (taken to be 6.5 unless otherwise specified), and ϵ_0 is the permittivity of free space. Summation of the potential drops across the structure results in

$$V_C = E_0 x_0 + E_n x_n \tag{9}$$

where V_C is the voltage of the metal electrode with respect to the silicon substrate.

Simultaneous solution of Eqs. (7)-(9) leads to

expressions for charge accumulation at the dielectric interface and current density in terms of the applied voltage and the dielectric thicknesses. In order to gain insight into the trends of the solution as a function of different parameters, the solution will be demonstrated at low temperatures. The similarity and relative simplicity of the oxide and the nitride current expressions in this temperature range [Eqs. (1) and (4)] allow for the derivation of closed-form expressions for current density and induced interface charge. Substitution of Eqs. (1) and (4) into Eq. (7) yields the current continuity in terms of the electric fields at low temperatures

$$C_0 E_0^2 \exp(-E_1 / |E_0|) = C_2 E_n^2 \exp(-E_2 / |E_n|) \tag{10}$$

To obtain a relatively simple solution, let us make the following assumptions (generally satisfied for the range

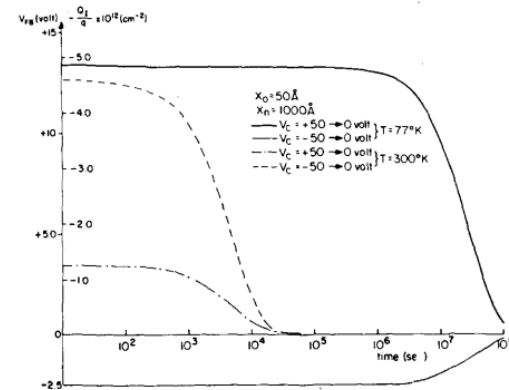


FIG. 8. Discharge of induced interface charge as a function of time for liquid nitrogen and room temperature.

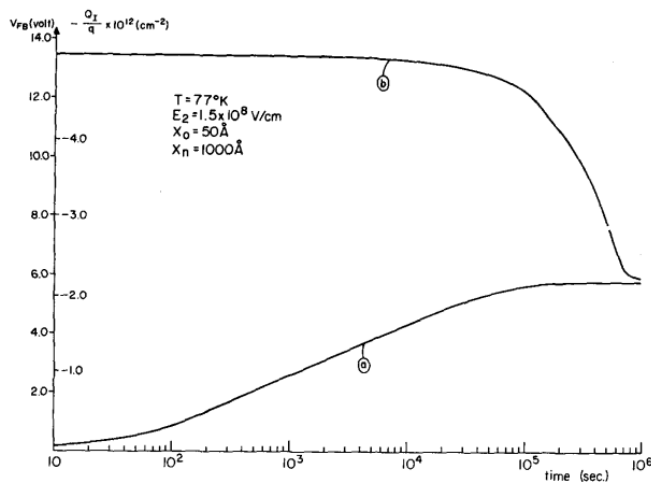


FIG. 9. Induced interface charge as a function of time for a given charging voltage ($V_C = +40 \text{ V}$) and two values of initial interface charge $[Q_I(0)/q]$. (a) $Q_I(0)/q = 0$. (b) $Q_I(0)/q = -5.0 \times 10^{12} \text{ cm}^{-2}$.

of practical interest):

- (a) $E_2 x_n \gg E_1 x_0, C_0 \geq C_2,$
- (b) $\ln(E_0^2/E_n^2) \ll \ln(C_0/C_2).$

Under these assumptions the simultaneous solutions of Eqs. (8)–(10) yield the following expression for the charge accumulated at the dielectric interface:

$$Q_I = \left[\frac{\epsilon_0 K_0 E_1}{\alpha |V_C| + x_0 E_1 + x_n E_2} - \frac{\epsilon_0 K_n E_2}{x_0 E_1 + x_n E_2} \right] V_C, \quad (11)$$

where

$$\alpha = \ln(C_0/C_2)$$

for positive V_C . For negative V_C

$$\alpha = \ln(C_0/-C_2),$$

and E_{1-} is substituted for E_1 . The electric fields are given by

$$E_0 = E_1 V_C / (\alpha |V_C| + x_0 E_1 + x_n E_2), \quad (12)$$

$$E_n = E_2 V_C / (x_0 E_1 + x_n E_2). \quad (13)$$

Substitution of the electric field in Eq. (10) results in the current-density expression

$$|J_0| = |J_n| = C_n \left[\frac{E_2 V_C}{x_0 E_1 + x_n E_2} \right]^2 \exp \left[-\frac{x_0 E_1 + x_n E_2}{|V_C|} \right]. \quad (14)$$

Solutions for the current density as a function of applied voltage for a typical structure, based on Eq. (14), are shown in Fig. 3. The results for both positive and negative charging voltages are compared with computer

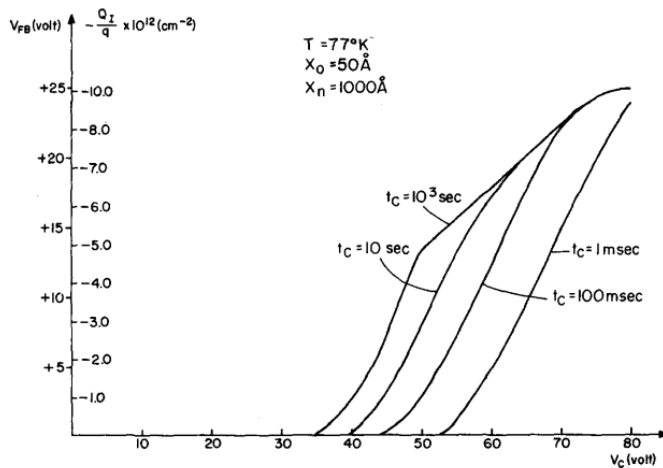
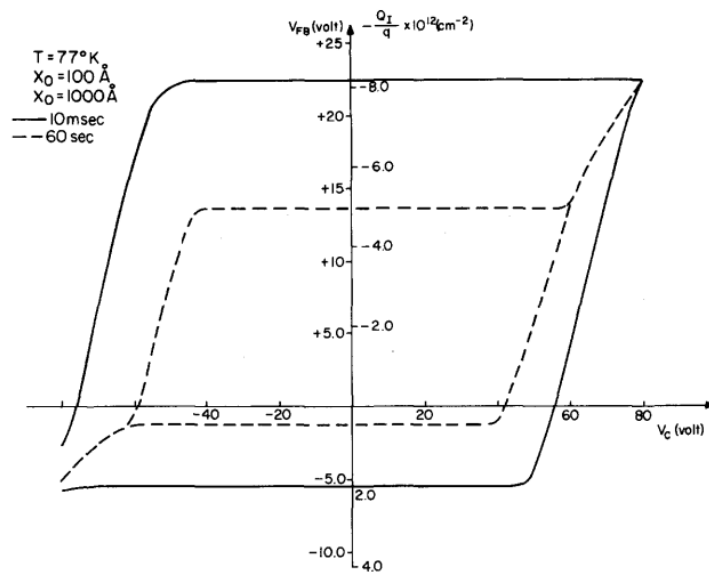


FIG. 10. Induced interface charge as a function of charging voltage for different values of charging time [$Q_I(0) = 0$].

Fig. 11. Hysteresis behavior of induced interface charge as a function of charging voltage for two values of charging time. $V_{C \min} = -80 \text{ V}$, $V_{C \max} = +80 \text{ V}$, $\Delta V_C = 5 \text{ V}$.



solutions based on Eqs. (7)–(9). Agreement between the solutions is good since the simplifying assumptions (a) and (b) are satisfied. Hence, Eq. (11) can be used to determine the accumulated charge dependence on the current-field characteristics and thickness of the dielectrics.

1. *Dependence of charge accumulation (Q_I) on the current-field characteristics and thickness of the dielectrics.* Equation (11) reflects the strong dependence of charge accumulation at the silicon nitride-silicon dioxide interface on the characteristic parameters (E_1 , E_2) of the current-field relationships [Eqs. (1) and (4)] for the dielectrics. For a fixed dielectric thickness ratio (x_0/x_n) and a given applied voltage, it predicts accumulation of either positive or negative charge at the interface, depending on the relative magnitude of these characteristic parameters. If we assume the initial nitride-oxide interface charge density to be zero and define the initial dielectric currents J_{00} , J_{n0} under the condition of no induced charge at the interface [$Q_I(V_C) = 0$]

$$\begin{aligned} J_{00} &= J_0(Q_I=0), \\ J_{n0} &= J_n(Q_I=0), \end{aligned} \quad (15)$$

then the polarity of the induced charge is determined by the difference between the initial current densities $J_{n0}(V_C) - J_{00}(V_C)$. The charge at the silicon dioxide-silicon interface does not affect the field distribution in the dielectrics. A fixed charge Q_{s0} at this interface will result in a constant shift of the flatband voltage equal to

$$\Delta V_{FB} = -(Q_{s0}/\epsilon_0) [(x_n/K_n) + (x_0/K_0)].$$

⁷ A. S. Grove, B. E. Deal, E. H. Snow, and C. T. Sah, *Solid-State Electron.* **8**, 145 (1965).

In the following theoretical calculation the silicon dioxide-silicon interface charge is assumed to be zero.

To illustrate the dependence of induced interface charge on the characteristic conduction parameters of the dielectrics, the accumulated charge is plotted as a function of charging voltage at low temperature for two values of E_2 (Fig. 4); negative charge corresponding to a positive flatband voltage $V_{FB} = -x_n Q_I / K_n \epsilon_0$ is plotted in the positive direction of the ordinate. At large voltages ($|V_C| > 40 \text{ V}$) Eq. (11) is in good agreement with computer solutions based on Eqs. (1), (2), and (7)–(9). A decrease in the value of E_2 , and consequently an increase in the value of J_{n0} with respect to J_{00} , leads to a more positive accumulated charge for positive applied voltages and to the opposite tendency for negative voltages. At low voltages the current in the nitride is dominated by the ohmic component J_{ns} [Eq. (5)], which is not taken into account in the derivation of Eq. (11). The deviation resulting from this increase in the nitride current is again in the direction of more positive charge for positive applied voltages.

So far we have assumed a fixed dielectric thickness ratio (x_0/x_n). A variation in this ratio will clearly lead to a modification of the electric field distribution. This in turn will result in a variation of the induced charge for a given charging voltage. Solutions including the ohmic component J_{ns} are shown in Fig. 5. In agreement with Eq. (11), an increase in the oxide thickness leads to a more positive interface charge for large positive charging voltages.

2. *Temperature dependence of Q_I .* So far we have restricted the discussion to low temperatures because of the relative simplicity of the current-field expression in

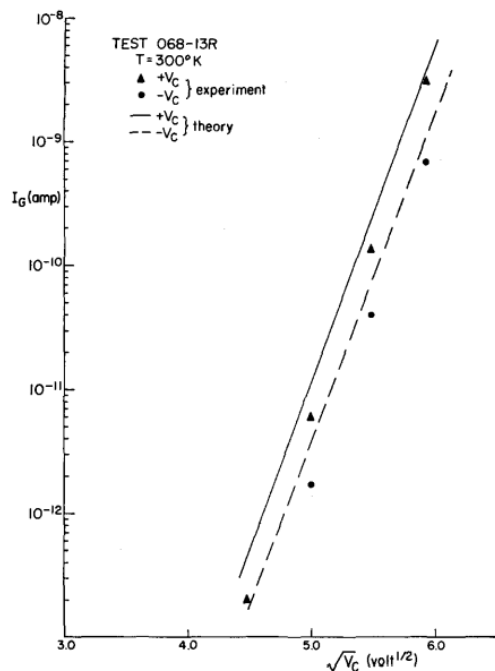


Fig. 12. Theory and experiment for the current as a function of $(V_G)^{1/2}$ at room temperature. $x_0=50 \text{ \AA}$, $x_n=530 \text{ \AA}$, area = $1.3 \times 10^{-4} \text{ cm}^2$.

this region. As the temperature is raised to room temperature and above, the change in relative magnitude of the current-density terms in Eqs. (1) and (2) will result in a modification of induced interface charge characteristics. At room temperature and above, the current density in the silicon nitride is dominated at moderate and high voltages by J_{n1} [Eq. (3)], and at low voltages by the ohmic component J_{ns} [Eq. (5)]. For silicon dioxide the temperature-dependent term in Eq. (1), which was negligible at low temperatures, also leads to an increase in current. However, the silicon nitride current increases more rapidly than the silicon dioxide current, due to the exponential temperature dependence of J_{n1} . The computer solution for induced charge as a function of applied charging voltage is shown in Fig. 6 and compared with the low-temperature solution for the same structure. As can be seen from a comparison of Figs. 4 and 6, the effect of higher temperature on the accumulated interface charge is in the same direction as the tendency observed for a lower value of E_2 . This similarity is expected since the net result of both increased temperature and a lower value of E_2 is an increase of current density in the silicon nitride with respect to the silicon-dioxide component for the same applied voltage.

Finally, it should be noted that the charge-transport model predicts the induced charge in the steady state to

be a single-valued function of the applied gate voltage. This implies that the MNOS device does not perform a memory function if operated under steady-state conditions. The reported charge storage characteristics^{1,2,8,9} are due to the transient behavior to be discussed next.

B. Transient Analysis

The rate of change in interface charge will be directly proportional to the difference in current-density values for both dielectrics at a given charging voltage and time

$$dQ_I(V_G, t)/dt = J_0(V_G, t) - J_n(V_G, t). \quad (16)$$

The initial condition for induced interface charge at the nitride-oxide interface is determined by the past history of the device. The computer solutions of Eq. (16) for a given MNOS structure at both low and room temperatures are shown in Figs. 7 and 8. The charging-time dependence for both voltage polarities ($V_G = \pm 50 \text{ V}$) is illustrated in Fig. 7. Note the significant decrease in charging response time at room temperature due to the increased current levels. The time response of charge decay after the charging voltage is removed ($V_G=0$) is illustrated in Fig. 8. Again, the room-

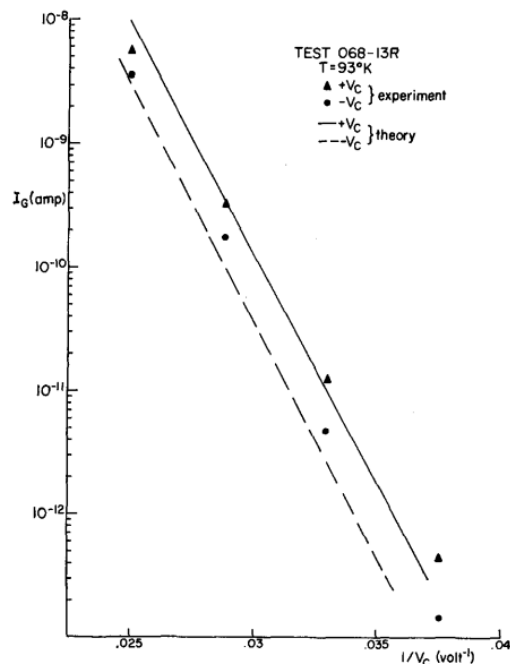


Fig. 13. Theory and experiment for the current as a function of $1/V_G$ at low temperature. $x_0=50 \text{ \AA}$, $x_n=530 \text{ \AA}$, area = $1.3 \times 10^{-4} \text{ cm}^2$.

⁸ J. T. Wallmark and J. H. Scott, Jr., "Switching and Storage Characteristics of MIS Memory Transistors," IEEE Int. Electron Devices Meeting, Washington, D.C., October 1968.

⁹ B. V. Keshavan and H. C. Lin, "MONOS Memory Element," IEEE Intern. Electron Devices Meeting, Washington, D.C., October 1968.

מאמר 1974

FAMOS—A NEW SEMICONDUCTOR CHARGE STORAGE DEVICE

DOV FROHMAN-BENTCHKOWSKY
INTEL Corporation, Santa Clara, California 95051, U.S.A.

(Received 13 August 1973; in revised form 15 October 1973)

Abstract—A new non-volatile charge storage device is described. The floating gate avalanche injection MOS (FAMOS) structure is a p -channel silicon gate field effect transistor in which no electric contact is made to the silicon gate. It combines the floating gate concept with avalanche injection of electrons from the surface depletion region of a p - n junction to yield reproducible charging characteristics with long term storage retention.

INTRODUCTION

In recent years a growing interest has been expressed in nonvolatile semiconductor memories. This interest stems from the need for high density, low cost semiconductor memory chips as well as an attempt to provide a substitute for the nonvolatile storage capability (retention of stored information without an external power source) of magnetic memories. Initial research efforts in this direction have centered in the area of semiconductor Read Only Memories (ROM's). Most available semiconductor ROM's are programmed permanently at the integrated circuit fabrication stage by a custom mask which defines the desired information pattern. As a result, program changes involve the generation of a new mechanical mask for every modified ROM pattern. In addition to being an expensive step, it also limits the flexibility of ROM applications because of the delay involved in the production process. Substitution of the delay in the generation and processing of a custom mask provides an ideal application area for nonvolatile semiconductor storage devices, since it does not impose a stringent requirement on the speed of electrical programming.

Most proposed non-volatile semiconductor storage devices rely on charge storage in a dielectric which forms part of the gate of an insulated gate field effect transistor. Feasibility has been demonstrated for MNOS (metal-nitride-oxide-silicon) memory devices[1,2] MAS (metal-aluminum-oxide-silicon) memory devices[3], and a dual gate MNOS memory device[4]. Difficulties in controlling the electrical characteristics of the storage dielectrics and additional fabrication steps required to achieve on-the-chip

decoding, have limited the realization of these approaches to undecoded memory arrays of up to 256 memory bits.

Recently, feasibility of the ovonic amorphous semiconductor memory device has been demonstrated by fabrication of an undecoded 256 bit memory array[5].

The recent introduction[6] of a novel semiconductor memory element—the FAMOS charge storage device, and its implementation in a fully decoded 2048 bit electrically programmable ROM[7], constitutes a significant advance in the state of the art of non-volatile semiconductor memories. It is the first available large capacity programmable ROM in which the information pattern is recorded electrically by way of a reversible change in memory device characteristics.

The Floating gate Avalanche injection MOS (FAMOS) memory device is essentially a p -channel silicon gate MOS insulated gate field effect device in which no electrical contact is made to the silicon gate. Operation of the FAMOS memory structure depends on charge transport to the floating gate by avalanche injection of electrons from either the source or drain p - n junctions.

The concept of an insulated gate field effect transistor with a floating gate as a non-volatile memory element was first advanced by Kahng and Sze[8]. Operation of the proposed structure was based on charge transport from the silicon substrate across a thin insulator layer ($\sim 50 \text{ \AA}$) to a floating metal electrode which is covered by a second insulator and an upper metal gate. Charge is stored in the floating metal gate in response to an applied voltage between the upper metal and the substrate. The charge transport mechanism

involves tunneling of electrons from the substrate through the thin oxide to the floating metal gate. The formation of a metal gate over a very thin dielectric layer is the major obstacle in practical realization of this proposed structure. The same concept evolved to the MNOS structure in which the floating metal gate was replaced by a layer of traps in the silicon nitride.

Nicollian *et al.* [9], reported that high current densities can be achieved in MOS capacitors by avalanche injection of electrons from a *p*-type silicon substrate. They observed a considerably lower current density level due to hole injection from an *n*-type substrate.

The FAMOS memory structure investigated in the present work combines the floating gate concept with avalanche injection of electrons to yield a new nonvolatile memory element.

DEVICE STRUCTURE AND CHARACTERISTICS

A cross-section of the Floating gate Avalanche injection MOS (FAMOS) structure is shown in Fig. 1. It is a *p*-channel silicon gate MOS insulated gate field effect device (made on 5–8 Ω cm *n*-type material of <111> orientation), in which no electrical contact is made to the silicon gate. The floating gate is formed by deposition of a poly-silicon layer over 1000 Å of thermal oxide and is isolated from the top surface by 1.0 μm of vapor deposited oxide. The FAMOS memory structure depends on charge transport to the floating gate by avalanche injection of electrons from a *p*-*n* junction. A junction voltage in excess of -30 V applied to a *p*-

channel FAMOS device will result in the onset of injection of high energy electrons from the *p*-*n* junction surface avalanche region to the floating silicon gate. Since the silicon gate is floating, the electron current through the oxide results in the accumulation of a negative charge on the gate. For a *p*-channel FAMOS device this negative charge will induce a conductive inversion layer connecting source and drain. The amount of charge transferred to the floating gate is a function of the amplitude and duration of the applied junction voltage. Once the applied junction voltage is removed, no discharge path is available for the accumulated electrons since the gate is surrounded by SiO₂, which is a very low conductivity dielectric. The accumulated negative charge on the gate will result in a change of device turn on voltage given by

$$\Delta V_T = - \left(\frac{Q_{G'} - Q_{G'}(0)}{C_0} \right)$$

where $Q_{G'}$ is the charge per unit area transferred to the floating gate, $Q_{G'}(0)$ is the initial charge on the floating gate, $C_0 = \epsilon_0 K_0 / X_0$ is the capacitance per unit area of the thermal oxide layer, ϵ_0 is the permittivity of free space, K_0 is the relative dielectric constant of silicon dioxide and X_0 is the thermal oxide layer thickness.

Device turn on voltage V_T is given by

$$V_T = V_{FB} + 2\phi_F - \frac{Q_B}{C_0} - \frac{Q_{G'}(0)}{C_0}$$

FLOATING GATE AVALANCHE INJECTION MOS
FAMOS

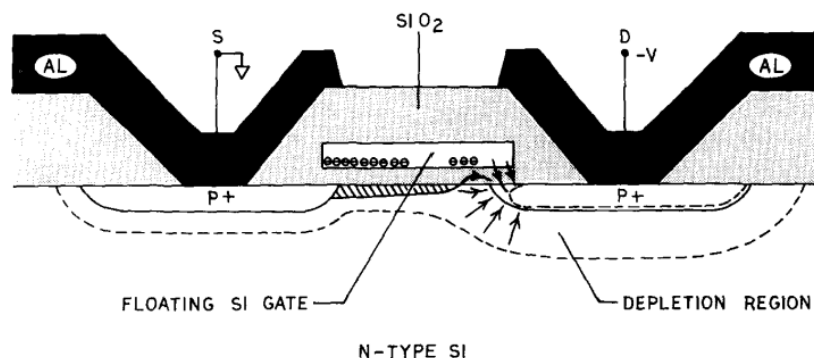


Fig. 1. Cross-section of FAMOS device under bias.

where Q_B is the charge within the substrate surface depletion region, active feedback voltage $V_{G'f}$ is given by

$$V_{FB} = \phi_{SS} - \frac{Q_{SS}}{C_0}$$

$$V_{G'f} = V_D \frac{C_{g'd}}{C_{g'd} + C_{g's} + C_{g'b}}$$

is the flat-band voltage, ϕ_{SS} is the polysilicon-silicon work function difference, Q_{SS} is the fixed charge at the silicon-silicon dioxide interface, and ϕ_F is the Fermi potential of the silicon substrate.

The presence or absence of charge on the floating gate can be sensed by measuring the conductance between the source and drain regions. The I_D - V_D characteristics of a charged and uncharged FAMOS device are shown in Fig. 2. Note that the uncharged FAMOS device conducts current for $V_D > -8.0$ V even though no charge was transferred to the floating gate. The current is due to capacitive voltage feedthrough from the drain terminal to the floating gate. An illustration of the interelectrode capacitances in the FAMOS structure is shown in Fig. 3. If a negative voltage is applied to the drain of the device with the source and substrate at ground potential, a negative charge will be capacitively coupled to the floating gate by means of a displacement current through $C_{g'd}$ (Fig. 3(b)). The value of the corresponding neg-

where $C_{g'b}$ is the series combination of $C_{g'}$ and C_b (Fig. 3(b)). Once this capacitively coupled voltage to the floating gate exceeds the turn on voltage V_T , a conductive path is established between source and drain resulting in the onset of current flow. The same capacitively coupled voltage is responsible for the deviation of the charged FAMOS device I - V characteristic (Fig. 2) from that of an MOS transistor with the same geometry and an applied gate voltage equivalent to the amount of charge on the floating gate of the FAMOS device.

To account for the capacitive feedback effect, the amount of voltage capacitively coupled to the floating gate was measured as a function of applied drain voltage under two limiting conditions, $V_{G'f} = Q_{G'}/C_0 = 0$ which corresponds to no electronic charge on the floating gate, and $V_{G'f} > V_D$ which defines a charging condition of the gate resulting in complete inversion between source and drain without pinch-off. The feedback voltage value $V_{G'f}$ was obtained by measuring the drain to source current flow for a given drain voltage on the FAMOS device and determining the equivalent

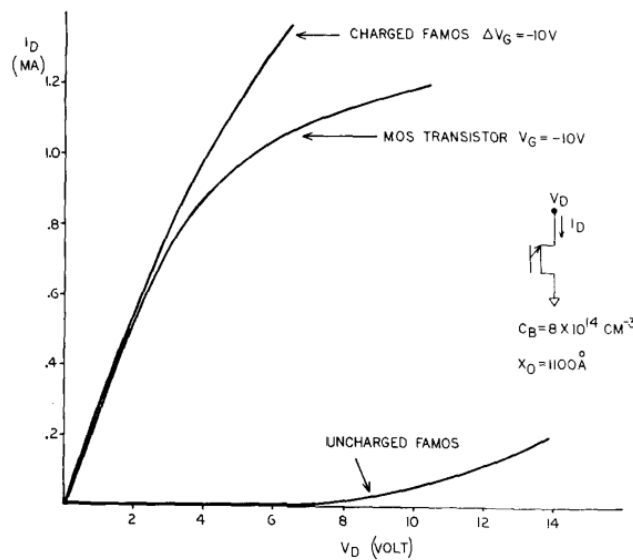


Fig. 2. I_D - V_D characteristics of charged and uncharged FAMOS device and the MOS transistor characteristic with a gate voltage corresponding to the amount of charge on the floating gate of the charged FAMOS device.

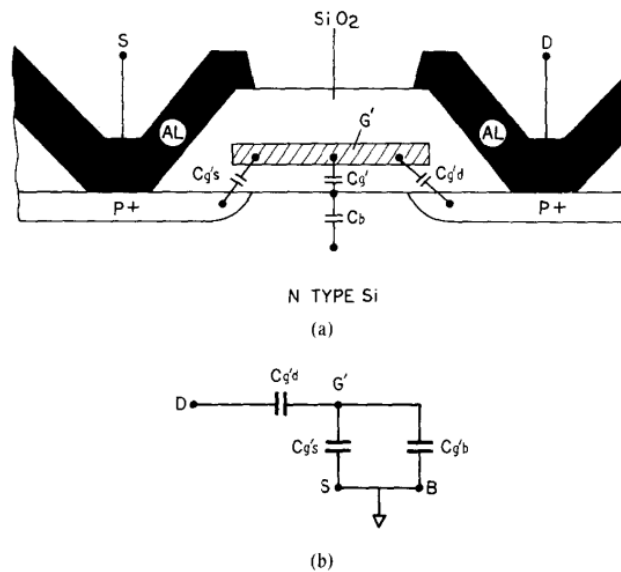


Fig. 3. (a) Cross-sectional view of the capacitances associated with the FAMOS structure. (b) Circuit representation of interelectrode capacitances.

gate voltage by comparison to an identical (process and geometry) MOS transistor. The two feedback voltage plots are shown in Fig. 4. The range of feedback factors $\Delta V_{G'}/\Delta V_D$ stems from the variation of interelectrode capacitances as a function of drain voltage. For the $V_{G'} > V_D$ shown in Fig. 5(a), the induced inversion layer extends from source to drain and the gate capacitance $C_{G'}$ splits approximately evenly between $C_{G's}$ and $C_{G'd}$ which is reflected in a high feedback factor $\Delta V_{G'}/\Delta V_D = 0.45$. The condition $V_{G'} > V_D$ was obtained by charging the FAMOS device to a maximum $V_{G'}$ -value of -22.0 V. In the case of $V_{G'} = 0$ (Fig. 5(b)) when the feedback voltage $V_{G'}$ is greater than the turn on voltage V_T , the FAMOS device I_D - V_D characteristic corresponds to that of an MOS transistor in saturation. The capacitance $C_{G'}$ will be partially diverted to the source ($C_{G's}$) through the pinched off inversion layer. The capacitance $C_{G'd}$ consists of a combination of the floating gate to drain overlap capacitance in parallel with the capacitive feedback through the pinched off surface depletion region C_{dep} (Fig. 5(b)). The linearity of the feedback plot for $V_{G'} > V_D$ is due to the presence of a continuous conductive path between source and drain such that the ratio of $C_{G'd}/C_{G's}$ is independent of V_D .

On the other hand, the reason for the apparent linearity of the $V_{G'} = 0$ plot over the measured range of V_D is more involved. While an increase

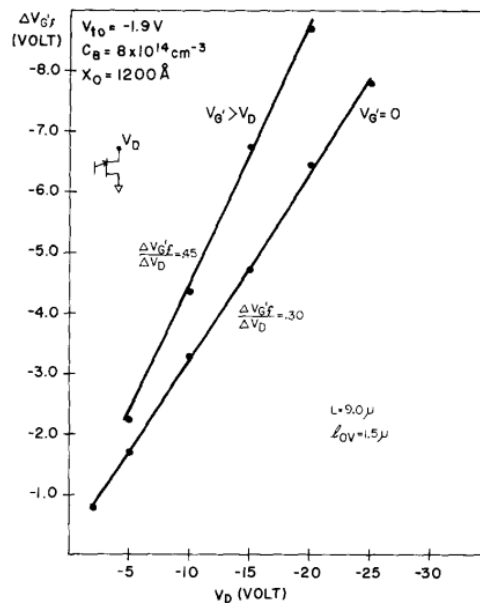


Fig. 4. Voltage feedback to the floating gate as a function of applied drain voltage.

in drain voltage is expected to increase the value of I_{dep} and thus affect the ratio of $C_{dep} + C_{G'd}/C_{G's}$ this tendency is countered by the effect of the electrical field in the oxide on I_{dep} [10]. This electric

FAMOS – A new semiconductor charge storage device

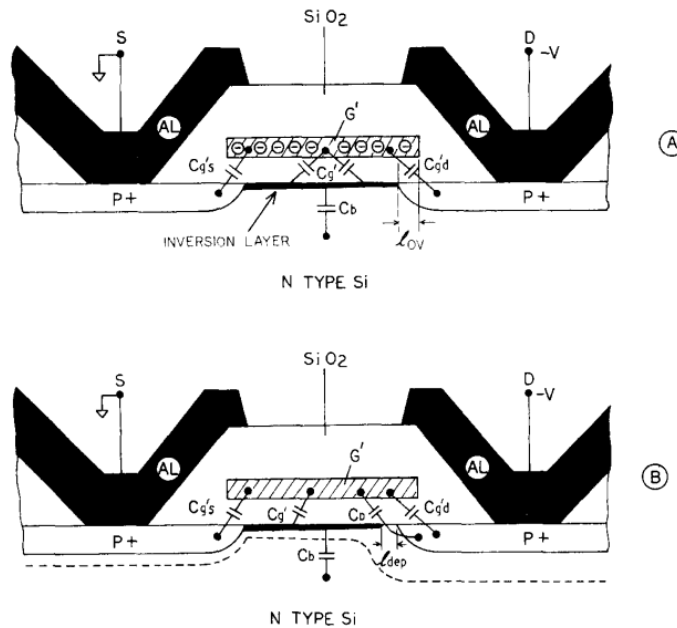


Fig. 5. Illustration of interelectrode capacitances for: (a) $V_{G'} > V_D$, (b) $V_{G'} = 0$.

field is a function of the potential difference $V_D - V_{G'}$ and tends to reduce the value of l_{dep} for increasing values of V_D . Hence the linearity of the $V_{G'} = 0$ plot can be traced to the opposing effect of the lateral field in the silicon and the vertical field in the oxide on the value of the drain surface depletion length l_{dep} .

Because of the pronounced effect of the feedback capacitance on device $I_D - V_D$ characteristics it is expected to play an important role in the charging behavior.

DEVICE OPERATION

Consider the FAMOS device cross section shown in Fig. 1. Initially all terminals are at a common ground potential. The floating silicon gate is not charged and assumes the common reference potential. As the drain voltage is increased negatively with respect to the source and substrate, a positive potential drop appears across the overlap region between the floating gate and the p^+ drain diffusion. This increasing positive voltage will tend to invert the heavily doped p^+ drain region; however, because of the lack of supply of minority carriers (due to the reverse biased condition of the drain $p-n$ junction) deep depletion will take place at the p^+ surface region of the drain near the SiO_2 interface. As the negative voltage on the

drain is increased further the electric field in the induced surface depletion region will reach a value at which an avalanche multiplication condition is established. Some of the high energy electrons generated in the avalanching surface depletion region acquire a sufficient energy to surmount the $Si-SiO_2$ energy barrier and be swept by the electric field in the SiO_2 toward the conductive floating silicon gate.

The avalanche injection process is illustrated in Fig. 6 in terms of the variation of the energy bands with distance in a direction normal to the p^+ drain region. Initially (Fig. 6(a)) the structure is in equilibrium. The applied negative voltage (Fig. 6(b)) results in impact ionization and emission of electrons from the p^+ surface depletion region over the $Si-SiO_2$ energy barrier to the floating silicon gate. Finally (Fig. 6(c)) when the applied voltage is removed, the electrons are trapped in a potential well, bounded by polysilicon-silicon dioxide energy barriers on both sides.

(a) *Charge accumulation*

A detailed investigation of avalanche injection in response to an applied voltage pulse to metal-oxide-silicon (MOS) capacitor structures was performed by Nicollian *et al.*[11].

The current densities observed for hot electron

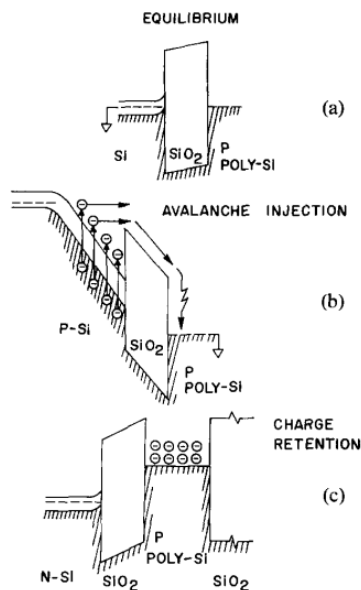


Fig. 6. Variation of the energy bands with distance in a direction normal to the p^+ drain region for a FAMOS device, (a) in equilibrium, (b) under avalanche injection, (c) after voltage removal.

emission through the SiO_2 layer of an MOS capacitor made on a p -type substrate were of the order of 10^{-4} A/cm^2 at an oxide field of $3.0 \times 10^6 \text{ V/cm}$. Considerably lower current density levels were observed for hole injection from an n -type substrate.

Measurements of current flow through the oxide of a silicon gate MOS transistor operated in the avalanche injection mode (Fig. 7) indicate that charge transport in the FAMOS structure can be attributed to the same hot electron emission mechanism observed in MOS capacitors. The current density data in Fig. 7 are comparable to those reported by Nicollian *et al.*[9] for a given field in the oxide. An exact comparison between the data is complicated by the gradual surface carrier concentration profile along the surface of the p^+ drain region. The two distinct plots in Fig. 7 correspond to the initial and saturation values of current density through the oxide. As pointed out by Poirier and Olivier[12], when the avalanche injection current is measured, a potential drift is observed which corresponds to trapping of negative charge in the oxide. As a result, for a given d.c. applied voltage, the electric field distribution in the oxide is time dependent, leading

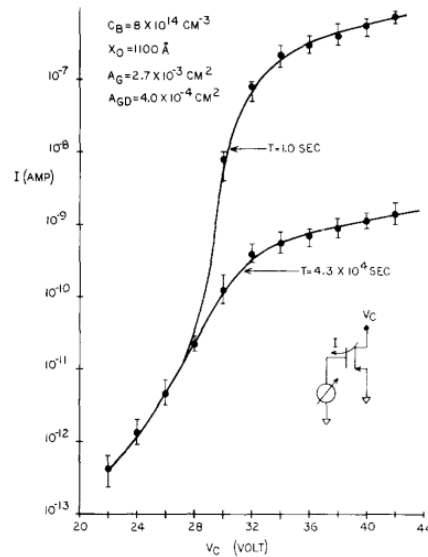


Fig. 7. Gate current as a function of applied drain voltage in an MOS transistor operating in the avalanche injection mode, for two values of applied voltage duration.

to different initial and saturation values of current density. The rapid increase in current density and the gradual saturation is due to avalanche multiplication followed by an avalanche breakdown condition. In the carrier multiplication region, an increase in applied voltage enhances the electric field in the surface depletion region which leads to an increase in the number of hot electrons injected over the Si-SiO_2 barrier to the silicon gate. When the surface field reaches the critical value for the onset of avalanche breakdown, any further increase in applied voltage is dropped across the thermal oxide, resulting in the saturation of current density. If under the same experimental conditions (Fig. 7) the silicon gate is floating, the avalanche injected current results in the accumulation of a negative electron charge on the gate. The amount of charge transferred to the floating gate as a function of the applied junction voltage pulse amplitude and width is shown in Fig. 8. To verify that the charging effect is due to electron storage on the silicon gate and not due to trapping in the oxide at either interface, the charging experiments were repeated with a conventional silicon gate MOS transistor in which the metal connecting to the gate was left floating. The charging behavior was identical to that of the FAMOS device, and the stored charge could be removed by connecting the gate to ground. A

FAMOS – A new semiconductor charge storage device

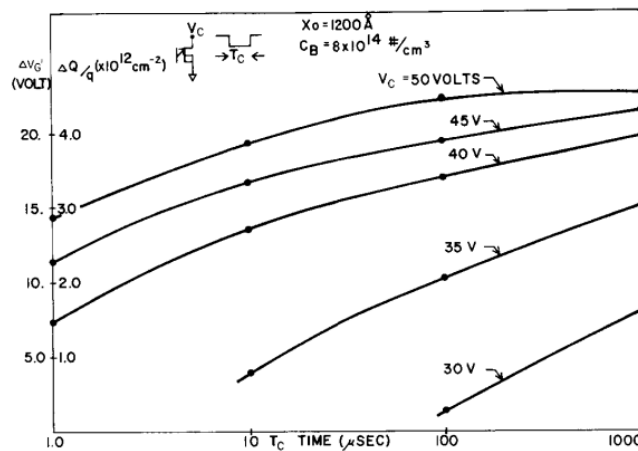


Fig. 8. Charge transferred to the floating gate of a FAMOS device as a function of applied charging voltage durations for different amplitudes of the applied voltage.

small residual charge of less than 10 per cent of the total accumulated charge was observed and attributed to trapping in the oxide[9,12]. The nature of this negative charge trapping will be interpreted further in the discussion of charge retention in the FAMOS structure. An attempt to correlate the current density data (Fig. 7) with the charge accumulated for a given applied voltage (Fig. 8) becomes difficult because of the different electric field distribution in the drain region of the MOS transistor and the FAMOS device. This difference in field distribution is caused by both the negative charge accumulation and capacitive feedback from the drain to the floating gate (detailed in the previous section). Both of these effects will reduce the potential difference between the floating gate and the drain, causing a decrease in the avalanche injected current with time. An accurate numerical account of these effects requires detailed knowledge of the time dependent field distribution in the deep depleted p^+ surface depletion region.

An additional effect not reflected in the current measurements in Fig. 7 is due to enhancement of carrier multiplication in the pinch off region of the FAMOS device by injection of carriers from the inversion layer. Such an enhancement of carrier multiplication affecting drain junction breakdown in an MOS transistor has been previously reported [13]. As shown in Fig. 9, gate current measurements under the condition $|V_G| > |V_T|$ (on the MOS transistor used in Fig. 7) indicate an increase of avalanche injection current which can be attributed to onset of current flow. An increase in

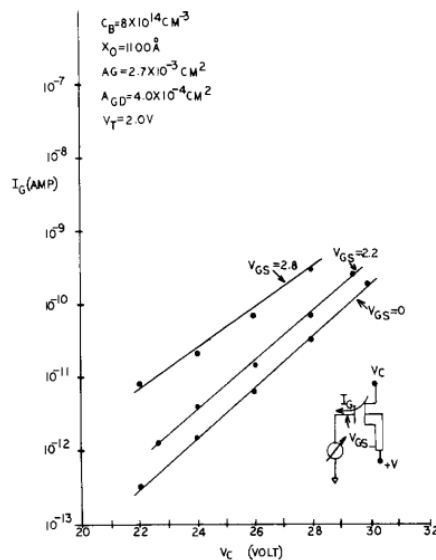


Fig. 9. Gate current as a function of applied drain voltage in an MOS transistor operating in the avalanche injection mode, for different values of gate to source voltage.

gate voltage to simulate current flow conditions during charging of the FAMOS device would lead to excessive currents through the MOS transistor because of the large gate area required to record the low level gate current.

Another aspect of interest concerning the charge accumulation data is the existence of a charging threshold. The value of this threshold was obtained from a measurement of charge saturation

as a function of applied charging voltage to the drain. Saturation was defined as charge accumulation after 12 hr of applied voltage. The results are shown in Fig. 10. From knowledge of the amount of charge accumulation and the feedback voltage to the gate (measured by the same method discussed with reference to Fig. 4), the saturation gate to drain voltage was determined as a function of applied voltage as shown in Fig. 11. The results indicate that saturation of charge transfer occurs at a constant gate to drain voltage of 8.0–9.0 V independent of the applied charging voltage. This value is within measurement error of the charging threshold voltage extrapolated from Fig. 10. The low charging threshold points to the possibility

that initial charging takes place by way of thermally generated electrons from the reverse biased surface depletion region.

(b) *Charge retention*

Once the applied charging voltage is removed, no discharge path is available for the accumulated electrons, since the gate is surrounded by thermal oxide which is a very low conductivity dielectric. The electric field in the structure after the removal of junction voltage is due only to the accumulated electron charge and is not sufficient to cause appreciable charge transport across the polysilicon–thermal oxide energy barrier. The maximum stored charge of 4×10^{12} No./cm² ($V_w = 50$ V, Fig. 8) results in an electric field of approximately 2×10^6 V/cm across the thermal oxide. Assuming current transport by Fowler–Nordheim emission from the poly-silicon gate into the oxide [14], the estimated discharge current (for a polysilicon–SiO₂ energy barrier of at least 3.2 eV) is of the order of 10^{-40} A/cm² at 300°K. Charge decay plots as a function of time for different initial charge values at 125 and 300°C are shown in Fig. 12. The rapid initial decay followed by a logarithmic time dependence cannot be explained by electron transport across the oxide. It has been shown [15,16] that an MOS structure subjected to a high electric field in the oxide by means of a negative bias applied to the gate at elevated temperature exhibits positive charge accumulation at the Si–SiO₂ interface. A charged FAMOS device incorporates the negative bias in the form of a negatively charged gate. Hence, one would

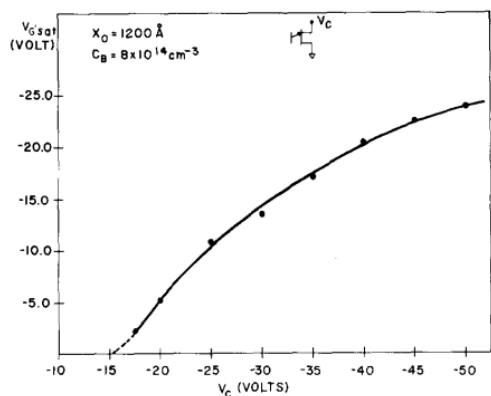


Fig. 10. The amount of charge transferred to the floating gate as a function of the steady state applied charging voltage.

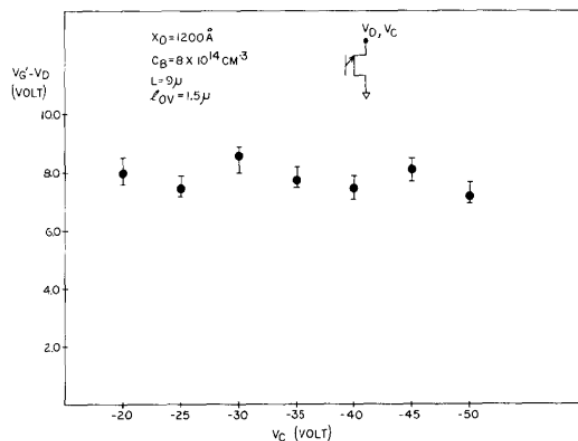


Fig. 11. Floating gate to drain potential difference as a function of the steady state applied charging voltage.

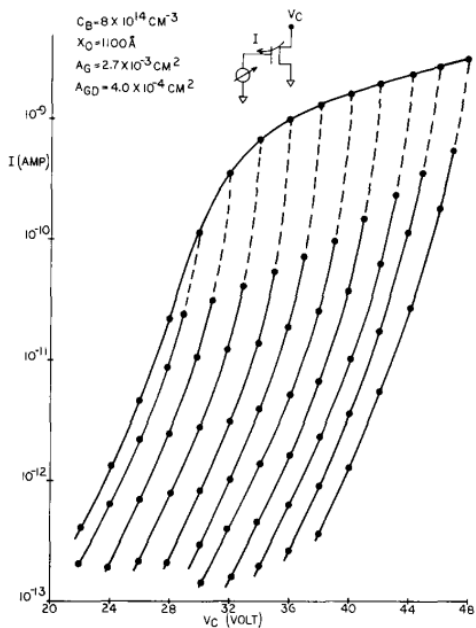


Fig. 14. Hysteresis behavior of gate current as a function of applied drain voltage in an MOS transistor operating in the avalanche injection mode. A_{GD} is the gate to drain overlap area.

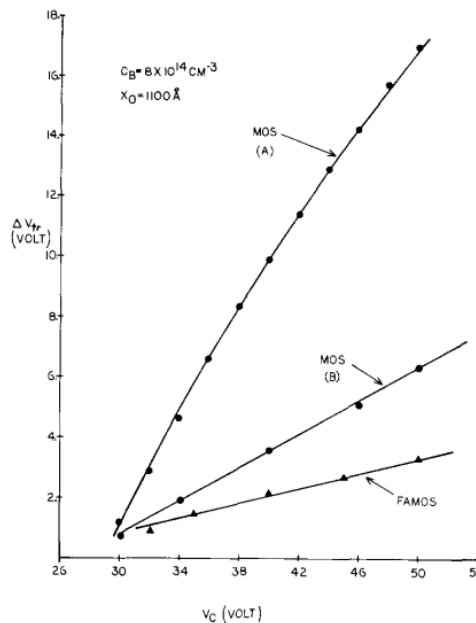


Fig. 15. Charge trapping in an MOS transistor operated in the avalanche injection mode. (a) before high temperature storage, (b) after 60 sec storage at 300°C. Initial charge loss of a FAMOS device stored 60 sec at 300°C is shown for comparison.

charge trapped and is plotted as a function of the steady state applied voltage V_C in Fig. 15. The ΔV_{tr} -values were obtained by determining the amount of voltage offset for each of the $I-V$ plots in Fig. 14 with respect to the (leftmost) steady state $I-V$ characteristic.

To evaluate the effect of high temperature storage on the trapped charge, the experiments leading to the $I-V$ plots in Fig. 14 were repeated with the following modification. After steady state was reached at a given applied voltage, the device was heated without bias for 60 sec at 300°C. The voltage shift ΔV_{tr} of the $I-V$ characteristics measured after storage is plotted as a function of steady state voltage V_C in Fig. 15(b). The lower values of ΔV_{tr} for a given V_C compared to plot A indicate a substantial loss of trapped charge during high temperature storage.

For qualitative comparison, the initial charge loss in a FAMOS device (as a function of the applied charging voltage) after 60 sec of storage at 300°C, is also plotted in Fig. 15. While the voltage and temperature dependence of both charge loss mechanisms is similar, the magnitudes are considerably different. If one assumes that the trapped charge is a function of the avalanche in-

jected current density, the magnitude of trapped charge in the FAMOS structure should be small compared to that of the MOS structure because of the lower gate to drain voltage drop for a given applied drain voltage (section 2(a)). However, a quantitative correlation between the charge loss mechanisms in both structures requires detailed knowledge of the field distribution in the avalanche injection region.

(c) Charge removal

To provide a useful memory function, it is desirable to be able to remove the charge from the floating gate. Since the gate is not electrically accessible, charge removal by means of an electrical pulse becomes difficult. The possibility of compensating the electron charge by injection of holes from the substrate is questionable due to the lack of evidence for substantial hole conduction through the oxide. However, the initial equilibrium condition of no electronic charge on the gate can be restored by illuminating the device with ultraviolet light or by exposure to X-ray radiation. The two charge removal processes are illustrated qualitatively in Fig. 16. Illumination with ultraviolet light (Fig. 16(a)) of the appropriate

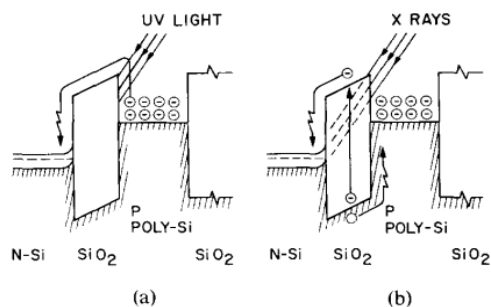


Fig. 16. Illustrations of charge removal by ultraviolet light and X-ray irradiation.

wavelength will impart sufficient photon energy to the stored electrons to excite them over the barrier with the aid of the built-in field in the oxide to the silicon substrate. If we assume a degenerate p^+ polysilicon gate, the energy barrier faced by the electrons is approximately 4.3 eV. Hence, an ultraviolet wavelength corresponding to a photon energy in excess of 4.3 eV will result in a discharging photo current from the floating gate to the substrate. A typical charge removal plot for a FAMOS device obtained by exposure to 2537 Å (4.9 eV) ultraviolet light is shown in Fig. 17.

Charge removal by X-ray radiation (Fig. 16(b)) involves the generation of hole electron pairs in the oxide by the ionizing radiation which has an energy greater than the band gap of SiO_2 (approx. 8.0 eV). A plot of charge removal as a function of radiation dose is shown in Fig. 18. The radiation dose required for complete charge removal, 5×10^4 rads, is comparable to the dose reported by

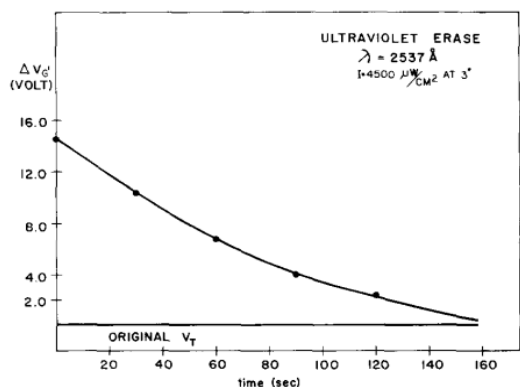


Fig. 17. The amount of charge removal by ultraviolet light as a function of time. I is the measured intensity of the light.

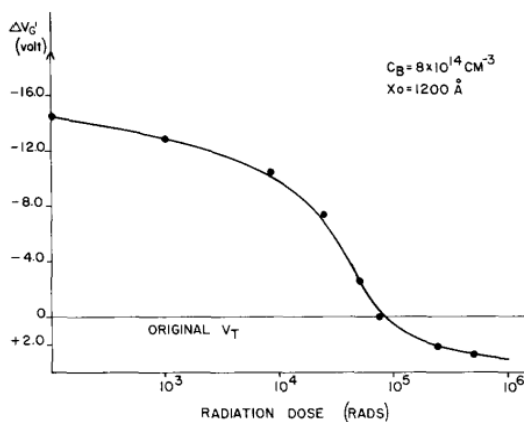


Fig. 18. Charge removal by X-ray irradiation as a function of radiation dose.

Snow *et al.* [17] for the onset of radiation induced turn on voltage shift in MOS transistors. As reported in the literature concerning the exposure of semiconductor devices to X-ray radiation, [17,18] the positive charge accumulation resulting from trapping in the oxide can be annealed at temperatures above 200°C. To evaluate the effect of exposure to ionizing radiation on the storage retention capability of the FAMOS device, the charge decay characteristics at 200°C were compared for irradiated (annealed at 200°C for 15 hr) and non-irradiated units charged to the same value. The results are shown in Fig. 19, and indicate that no significant difference in retention characteristics can be discerned.

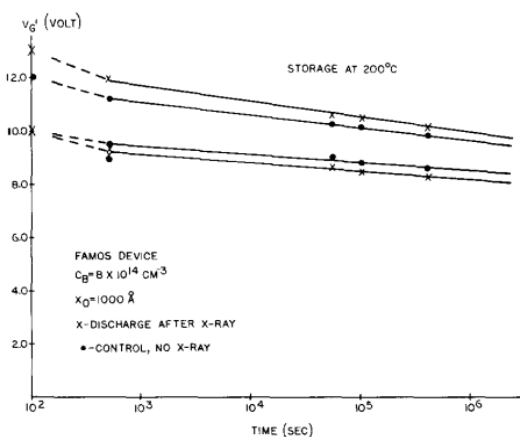


Fig. 19. Comparison of charge retention for irradiated FAMOS devices annealed 15 hr at 200°C with non-irradiated units.

DEVICE APPLICATIONS

A non-volatile charge storage device with long term retention and charge removal capability has many potential applications in the areas of digital and analog circuits. Semiconductor memories, photo-sensitive storage arrays, depletion mode load devices in digital circuits and analog storage circuits are a few specific examples. However, the most widespread potential application is in the area of semiconductor programmable read only memories. Fabrication of a monolithic 2048 bit fully decoded electrically programmable read only memory with FAMOS devices in the basic memory cell has been previously reported[7]. A photomicrograph of the integrated circuit memory chip is shown in Fig. 20. While most of the design aspects of the memory chip relate to circuit design techniques detailed in another publication[19], some of the considerations involving the memory cell deserve elaboration from a device standpoint. Two possible ways of incorporating a FAMOS device in a programmable read only memory cell are shown in Fig. 21. Since the storage device (T_1) is a two terminal device, an additional isolation MOS transistor (T_2) is required to avoid erroneous interpretation of information due to coupling between normally 'on' (charged) FAMOS devices in the memory array. In both schemes shown in Fig. 21, programming of the storage device is accomplished by a coincidence negative voltage applied to the X and Y select lines, resulting in avalanche injection and retention of charge on the floating gate of the storage device. Despite the apparent symmetry of the two schemes (Fig. 21) they differ in their programming characteristics. Configuration I (Fig. 21(a)) for the same T_2 device size results in

a lower voltage at the drain terminal of the FAMOS device for a given programming voltage V_p , due to the lower conductance of T_2 under this biasing configuration. Consequently, the amount of charge transferred to the floating gate for a given programming pulse amplitude and width is less for Configuration I compared with that of Configuration II.

Proper operation of the memory requires the capability to inhibit the programming of a non-selected bit. In Configuration I (Fig. 21(a)) this is simply accomplished by the lack of a Y -select signal or the absence of an X -select excitation, which inhibits the transfer of the programming pulse to the drain of the FAMOS device. On the other hand, in Configuration II (Fig. 21(b)) the inhibit mechanism in the case of no X -select signal is more complicated. If the X -select line is grounded while a negative voltage is present on the Y -select line, the drain of the FAMOS device (Fig. 21(b)) is still subjected to a high negative voltage, however, no avalanche injection will occur in this case because the source of the storage device is floating (T_2 is 'off'). The absence of avalanche injection can be understood with the aid of the capacitive feedback considerations discussed in Section I. Under the programming condition (T_2 'on') the capacitance distribution in the FAMOS device is similar to that illustrated in Fig. 5(b). The feedback factor $\Delta V_{Gf}/\Delta V_D$ is of the order of 0.3. Hence the value of V_{Gf} for a typical drain voltage of -40 V during programming is -12 V. The potential difference $V_{Gf} - V_D$ is substantially above the charging threshold shown in Fig. 11. However, if the source is floating (in the inhibit mode) the shunt capacitance to ground is very small since C_{gs} through the negatively

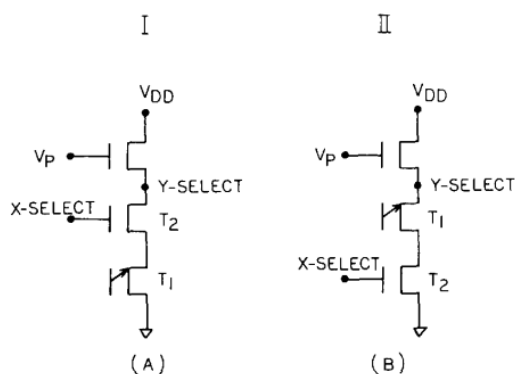


Fig. 21. Two configurations for incorporating FAMOS device in a programmable memory cell.

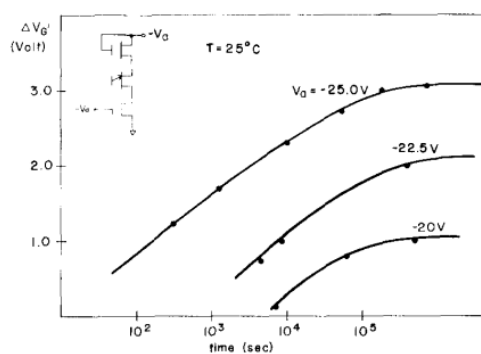


Fig. 22. Charge accumulation as a function of time in a memory cell for different values of applied READ voltage.

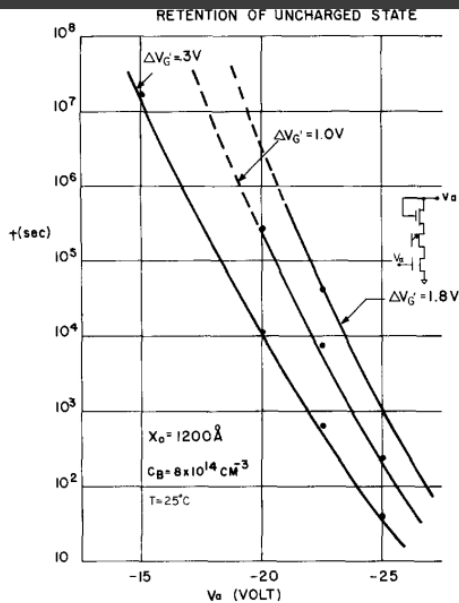


Fig. 23. Retention of the uncharged state in a memory cell as a function of the applied READ voltage.

rising voltage on the floating source provides additional feedback to the gate. As a result, the feedback factor can be as high as 0.8–0.9. $V_{G'} - V_D$ becomes smaller than the charging threshold and no avalanche injection occurs.

Once a memory cell has been programmed, it can be interrogated in the READ mode by coincidence selection (X, Y) with a lower negative voltage (-15 V). Hence the only difference between selection in the PROGRAM mode and the READ mode is the applied voltage level -50 and -15 V, respectively. This brings up the question of whether an uncharged memory cell can be slowly charged by repeated selection in the READ mode. The memory cell in configuration II (Fig. 21(b)) is particularly susceptible because of its advantage in programmability.

Parasitic charging in this configuration was measured for different voltages as a function of time. The results shown in Fig. 22 indicate charging characteristics followed by saturation similar to those shown in Fig. 8. Computation of the saturation value $V_{G'} - V_D$ yields 7.5 V for all three voltages, which is in good agreement with the

saturation value obtained from Fig. 11. To estimate the long term effect of parasitic charging on the retention of the uncharged memory state under READ mode operating conditions ($V_a = -15$ V), the time dependence is plotted in Fig. 23 as a function of applied voltage for different values of parasitic charge accumulation. A voltage shift of $\Delta V_{G'} = 1.0$ V at $V_a = -15$ V will occur at an extrapolated time of at least 10 years. Hence parasitic charging does not present a practical reliability problem for memory cell operation.

Acknowledgements—The author wishes to thank L. L. Vadasz, A. S. Grove and G. E. Moore for many helpful discussions; T. A. Rowe and G. Pasco for fabrication of the memory devices; and G. Greenwood for his assistance with testing and instrumentation.

REFERENCES

1. H. A. Wegener, A. T. Lincoln, H. C. Pao, M. R. O'Connell and R. E. Oleksiak, *Int. Electron Devices Meeting* Washington, D. C. (1969).
2. D. Frohman-Bentchkowsky, *Proc. IEEE Letts* **57**, 1190 (1969).
3. S. Nakanuma, T. Tsujuda, R. Igarachi, K. Cnoda, T. Wada and M. Makagiri, *ISSCC Digest of Technical Papers*, p. 68 (1970).
4. H. G. Hill and T. N. Toombs, *Solid-St. Electron.* **12**, 981 (1969).
5. R. G. Neale, D. L. Nelson and G. E. Moore, *Electronics*, p. 56 (1970).
6. D. Frohman-Bentchkowsky, *Appl. Phys. Letts* **18**, 332 (1971).
7. D. Frohman-Bentchkowsky, *ISSCC Digest of Technical Papers*, p. 80 (1971).
8. D. Kahng and S. M. Sze, *Bell Syst. Tech. J.* **46**, 1288 (1967).
9. E. H. Nicollian, A. Goetzberger and C. N. Berglund, *Appl. Phys. Letts* **15**, 174 (1969).
10. D. Frohman-Bentchkowsky and A. S. Grove, *IEEE Trans. Electron Devices*, **ED-16**, 108 (1969).
11. E. H. Nicollian and C. N. Berglund, *J. appl. Phys.* **41**, 3052 (1970).
12. R. Poirier and J. Olivier, *Appl. Phys. Letts* **15**, (1969).
13. H. Martinot and P. Rossel, *Electron. Letts* **7**, 118 (1971).
14. M. Lenzlinger and E. H. Snow, *J. appl. Phys.* **40**, 278 (1969).
15. B. E. Deal, M. Sklar, A. S. Grove and E. H. Snow, *J. Electrochem. Soc., Solid State Sci.* **114**, 266 (1967).
16. S. R. Hofstein, *Solid-St. Electron.* **10**, 657 (1967).
17. E. H. Snow, A. S. Grove and D. J. Fitzgerald, *Proc. IEEE*, **55**, 1168 (1967).
18. C. W. Gwyn, *J. appl. Phys.* **40**, 4886 (1969).
19. D. Frohman-Bentchkowsky, *IEEE Solid State Circuits J.* (1971).

- Frohman-Bentchkowsky, D. & Lenzlinger, M. (1969). Charge Transport and Storage in Metal–Nitride–Oxide–Silicon (MNOS) Structures. *Journal of Applied Physics*, 40(8), 3307–3319
- Frohman-Bentchkowsky, D. (1970). The Metal-Nitride-Oxide-Silicon (MNOS) Transistor — Characteristics and Applications. *Proceedings of the IEEE*, (8)58, 1219–1207
- Frohman-Bentchkowsky, D. (1971). Memory Behavior in a Floating-Gate Avalanche-Injection MOS (FAMOS) Structure. *Applied Physics Letters*, 18(8), 334–332
- Lenzlinger, M. & Snow, E.H. (1969). Fowler–Nordheim Tunneling into Thermally Grown SiO₂. *Journal of Applied Physics*, 40(1), 278–283
- Masuoka, F. et al. (1984). A New Flash EEPROM Cell Using Triple Polysilicon Technology. *IEEE IEDM Technical Digest*, pp. 464–467
- Masuoka, F. et al. (1987). A 256K Flash EEPROM Using Triple Polysilicon Technology. *IEEE JSSC*, 22(4), 548–552
- Bez, R. et al. (2003). Introduction to Flash Memory. *Proceedings of the IEEE*, (4)91, 502–489
- Kim, K. (2010). Future Memory Technology Including Emerging New Memories. *ESSDERC*, pp. 3–16
- Micheloni, R., Crippa, L. & Marelli, A. (2010). *Inside NAND Flash Memories*. Springer
- Nicollian, E.H. & Brews, J.R. (1982). *MOS Physics and Technology*. Wiley
- Sze, S.M. (1981). *Physics of Semiconductor Devices* (2nd ed.). Wiley