Decision-theoretic modeling of early life failures in semiconductor manufacturing

Jürgen Pilz$^1$   Daniel Kurz$^1$   Horst Lewitschnig$^2$

$^1$Institut für Statistik, Universität Klagenfurt
Universitätsstr. 65-67, 9020 Klagenfurt, Austria
juergen.pilz@aau.at
daniel.kurz@aau.at

$^2$Infineon Austria
horst.lewitschnig@aau.at

GDRR 2013
July 8-10, 2013 / Kinsale, Ireland
1. Introduction

2. Interval estimation for early life failure probabilities
   - 2.1 Clopper-Pearson interval estimation
   - 2.2 Bayesian equal-tail interval for $p$

3. Assessing ppm-levels using CM’s
   - 3.1 Single CM failure probability model
   - 3.2 Multiple CM failure model
   - 3.3 CM’s with uncertain effectivenesses

4. Decision-theoretical formulation of the CM failure probability model

5. Bayes decisions and application of the CM failure model

6. Bayesian assessment of Weibull early life failure distributions

7. Bibliography
1. Introduction

Key issue in semiconductor manufacturing:

Reliability

most commonly applied failure screening technique: **Burn-in-study**, especially in safety-critical applications

**Basis:** bathtub curve describing hazard rate
Testing under accelerated stress conditions  
(increased temperature & voltage stress)

**Burn-in:** independently selected number of devices is investigated for early failures

**Model** for early failures: **Weibull** distribution $Wb(a, b), b < 1$.

Current ppm-requirement: 21ppm (Infineon Technologies Villach, Austria)

Burn-in schemes different for logic and power devices. Here we focus on **power devices**.

Reasons for early failure: oxide particles, metallization defects,...
**Problem:** only very few failures

⇒ it’s rarely possible to efficiently fit a Weibull DFR distribution to burn-in data.

**Way out:** prove that early life failure probability $p \in$ target confidence area

Burn-in read-outs at discrete time points $t_1, t_2, t_3$

Report statistics: $k_j = \# \text{ failures in } (t_{j-1}, t_j]$

$$j = 1, 2, 3; \quad t_0 = 0$$
Goal: $P$ (early life failure after $t_3$ hours) $\leq 21 ppm$

Successful burn-in: requires $k = k_1 + k_2 + k_3 = 0$

(Zero defect strategy)

Usually: Burn-in is re-started whenever a failure occurs

Current standard: introduction of countermeasures (CM)
(ink out, design measures, optical inspection, ...)
to reduce the failure probability $p$

Our aim:

- development of a statistical model for taking account of CM’s
- avoid re-start of burn-in by planning additional number of items to be burnt for zero defects.
2. Interval estimation for early life failure probabilities

$n$ independently selected devices are stressed

\[ X_i = \begin{cases} 
0 & \text{if device } i \text{ passes the burn-in} \\ 
1 & \text{if device } i \text{ fails within burn-in} 
\end{cases} \]

\[ X = \sum_{i=1}^{n} X_i \sim Bi(n, p) \]

\[ x = (x_1, \ldots, x_n) \in \{0, 1\}^n; \quad k = x^T x \in \{0, 1, \ldots, n\} \]

\[ = \# \text{ failures} \]
2.1 Clopper-Pearson interval estimation

\[ l_{CP} = (\hat{p}_l, \hat{p}_u) \quad \text{where} \]

\[ P(X \geq k|\hat{p}_l) = \alpha/2 \quad \text{and} \]

\[ P(X \leq k|\hat{p}_u) = \alpha/2 \]

To obtain \( \hat{p}_l \) and \( \hat{p}_u \), we use the well-known relationship with the Beta distribution

\[ \hat{p}_l = F_{Z_l}^{-1}(\alpha/2) \quad \text{with} \quad Z_l \sim Be(k, n - k + 1) \]

\[ \hat{p}_u = F_{Z_u}^{-1}(1 - \alpha/2) \quad \text{with} \quad Z_u \sim Be(k + 1, n - k) \]

90% one-sided interval \( l_p = [0, \hat{p}_u]; \quad \alpha/2 = 0.1 \)
2.2 Bayesian equal-tail interval for $p$

In a Bayesian framework, this relationship comes in naturally observing that the conjugate prior for $p$ is the Beta distribution:

$$ p \sim Be(a, b); \quad a, b > 0 $$

$$ \Rightarrow f(p|\mathbf{x}) \propto l(p; \mathbf{x})f(p) = p^{a+k-1}(1 - p)^{b+n-k-1} $$

i.e. $p|\mathbf{x} \sim Be(a^* = a + k, b^* = b + n - k)$
Bayesian equal-tail credible interval

\[ C_e = (\hat{p}_l^*, \hat{p}_u^*) \text{ where } \hat{p}_l^* = F_{p|\mathbf{x}}^{-1}(\alpha/2), \hat{p}_u^* = F_{p|\mathbf{x}}^{-1}(1 - \alpha/2) \]

Jeffreys’ prior: \( a = b = 1/2 \)

Choosing \( a = 1, b = 0 \) we have

\[ p|\mathbf{x} = Be(k + 1, n - k) \]

\[ \hat{p}_u^* = \hat{p}_u \]

Concidence of one-sided Bayesian interval with Clopper-Pearson interval
Repair is impossible for semiconductor devices; they either pass or fail within the burn-in.

If a burn-in related failure occurs, then a CM is introduced (optical inspection, process improvement, ...) aiming to reduce $p$ to $\pi \leq p$.

**Crucial:** Experts assess the CM’s effectiveness $\vartheta \in [0, 1]$

\[ \vartheta = \text{probability of correcting the failure.} \]
3.1 Single CM failure probability model

Consider $k$ failures for which a single CM with effectiveness $\vartheta \in [0, 1]$ is implemented in the process.

**Interpretation:** There is a likelihood $\xi_j$ that $j \leq k$ failures would have occurred or, equivalently, $k - j$ failures would have been corrected if the CM would have already been introduced before the burn-in study.

Let $K_l = \begin{cases} 1 & \text{if failure } l \text{ is corrected} \\ 0 & \text{else} \end{cases}$

Clearly: $K = \sum_{l=1}^{k} K_l \sim Bi(k, \vartheta)$

$\downarrow$

unknown number of failures that would have been caught by the CM

$\Rightarrow$ (*) $\xi_j = P(K = k - j); \ j \in \{0, \ldots, k\}$
Clopper-Pearson model for single CM

after the CM: $X' \sim Bi(n, \pi)$

Weighting of Clopper-Pearson upper limits according to (*) leads to assessing $\hat{\pi}$ as

$$\sum_{j=0}^{k} \xi_j P(X' \leq j | \hat{\pi}) = \alpha$$

Equivalently: using $P(X' \leq j | \pi) = 1 - P(Z_j < \pi)$

with $Z_j \sim Be(j + 1, n - j); \ j = 0, \ldots, k$

$\Rightarrow \hat{\pi} = F_{Z'}^{-1}(1 - \alpha) = (1 - \alpha)$-quantile of

$Z' \sim \sum_{j=0}^{k} \xi_j Be(j + 1, n - j)$ Beta mixture
Bayesian model for single CM

prior $\pi \sim Be(a, b)$

actual number of failures after CM introduction is $k - K$ and is unknown

Therefore consider the preposterior:

$$\Xi := E[\pi | k - K] = \sum_{j=0}^{k} \xi_j(\pi | j) \sim \sum_{j=0}^{k} \xi_j Be(a + j, b + n - j)$$

$$\rightarrow \hat{\pi}^* = F_\Xi^{-1}(1 - \alpha) = (1 - \alpha) \text{-quantile of the mixture distribution } \Xi.$$ 

Again: $\hat{\pi}^* = \hat{\pi}$ for the prior $\pi \sim Be(1, 0)$

Setting $\vartheta = 0$ (no CM is implemented) we arrive at the classical estimation models.
3.2 Multiple CM failure model

now consider $r \geq 1$ different CM’s and denote $\vartheta = (\vartheta_1, \ldots, \vartheta_r) =$ vector of effectivenesses; $r \leq k$

$k = (k_1, \ldots, k_r)$; $k_i =$ # failures tackled by CM$_i$

$$\sum_{i=1}^{r} k_i = k$$

Now: $K = \sum_{l=1}^{k} K_l \sim GBi(k, \vartheta_k)$ generalized binomial, where

$$\vartheta_k = (\underbrace{\vartheta_1, \ldots, \vartheta_1}_{k_1 \text{ times}}, \underbrace{\vartheta_2, \ldots, \vartheta_2}_{k_2 \text{ times}}, \ldots, \underbrace{\vartheta_r, \ldots, \vartheta_r}_{k_r \text{ times}})$$

We have developed an efficient method for computing generalized binomial probabilities employing sequential convolution.
3.3 CM’s with uncertain effectivenesses

So far: $\vartheta_i, i = 1, \ldots, r$; were fixed

often: process experts are uncertain about the effectivenesses of the applied CM’s.

a) Beta-Binomial model for a single uncertain effectiveness, $r = 1$

\[ \vartheta \sim Be(u, v) \]

\[ K | \vartheta \sim Bi(k, \vartheta) \]

\[ \Rightarrow \xi_j = P(K = k - j) = \int_0^1 P(K = k - j | \vartheta) f(\vartheta) d\vartheta \]

\[ = \binom{k}{k - j} \frac{\Gamma(u+k-j)\Gamma(v+j)}{\Gamma(u+k+v)} \frac{\Gamma(u+v)}{\Gamma(u)\Gamma(v)} \]

\[ K \sim BeBi(k, u, v) \text{ Beta-Binomial} \]
b) Generalized Beta-Binomial model for more than a single uncertain effectiveness

\[ K | \vartheta \sim GBi(k, \vartheta_k) \]

\[ \vartheta_i \sim Be(u_i, v_i); \quad i = 1, \ldots, r \]

\[ \Rightarrow P(K = k - j) = \int_{[0,1]^r} P(K = k - j | \vartheta) f(\vartheta) d\vartheta \]

\[ K \sim GB\text{e}Bi(k, u_1, \ldots, u_r, v_1, \ldots, v_r, k_1, \ldots, k_r) \]

no closed form solution available,

MC-integration
4. Decision-theoretical formulation of the CM failure probability model

**Parameter space:** $p \in \Theta = [0, 1]$

after implementing CM’s: $\pi \in \Theta' = \Theta = [0, 1]$ with $\pi \leq p$

**Action space:** without CM’s $a = \hat{p} \in A = [0, 1]$

after incorporating CM’s: $a' = \hat{\pi} \in A' = A = [0, 1]$

**Sample space** of Burn-in data: $X|p \sim Bi(n, p)$

$x = (x_1, \ldots, x_n) \in \mathcal{X} = \{0, 1\}^n$

can be sufficiently described by

$\mathcal{T} = \{x^T x : x \in \mathcal{X}\} = \{0, 1, \ldots, n\}$
after implementing CM’s: we simulate failure scenarios $j \in T$, based on the observed $k \in T; \ 0 \leq j \leq k$; as outcomes, which would have possibly occurred if we would have introduced the CM’s already before the burn-in.

To these scenarios we attach prob’s $\xi_j$ (assessed wrt. the CM’s effectivenesses)

Assessment of the $\xi_j$: for single CM by means of $Bi(k, \vartheta)$, i.e. simulation depends on $k \in T$ and $\vartheta \in [0,1]$. 
in case of $r \leq k$ different CM’s:

$\xi_j$ determined by $GBi(k, \vartheta_k)$ where $k = (k_1, \ldots, k_r) \in \mathcal{K}$ reports the number of failures $k_j$ tackled by $CM_i$; $i = 1, \ldots, r$. There are

$$|\mathcal{K}| = \binom{r + k - 1}{k}$$

different vectors $k$

Simulations depend on observed $k \in \mathcal{T}, \vartheta \in [0, 1]^r$ and $k$
Decision functions

$d : \mathcal{T} \rightarrow A$ and $d(k) = \hat{\rho}$ ppm-level estimator extension in the CM decision framework

Single CM case: $d' : \mathcal{T} \times [0, 1] \rightarrow A'$ with $d'(k, \vartheta) = \hat{\pi} \in A'$

Multiple CM case: $d' : \mathcal{T} \times [0, 1]^r \times \mathcal{K} \rightarrow A'$

with $d'(k, \vartheta, k) = \hat{\pi} \in A'$
under-estimation of $p$ and $\pi$, resp., is more critical than over-estimation

propose asymmetric linear loss, i.e.

$$L(p, d(k) = \hat{p}) = \begin{cases} l_1(p - \hat{p}) & \text{if } \hat{p} \leq p \\ l_2(\hat{p} - p) & \text{if } \hat{p} > p \end{cases}$$

for the other cases: replace $p$ and $d$ by $\pi$ and $d'$, respectively.
Risk function

in the most general case of multiple CM we have

\[ R((\pi, \vartheta), d') = \sum_{k=0}^{n} \sum_{i=1}^{|\mathcal{K}|} L(\pi, d'(k, \vartheta, k_i)) \]

\[ \times \sum_{j=0}^{k} \xi_{ij} P(X' = j | \pi) \]

where \( \xi_{ij} = P(K = k - j) \) with \( K \sim GBi(k_i, \vartheta_{k_i}) \)

\[ i = 1, \ldots, |\mathcal{K}|; \quad j = 0, \ldots, k \]
Bayes decisions and application of the CM failure model

consider only CM decision framework with a single CM

need to specify a prior $f(\pi)$

Bayes optimal solution minimizes the preposterior expected loss: with

$\pi \sim Be(a, b)$ we obtain the preposterior distribution as Beta mixture

$$
\pi|k, \vartheta \sim \sum_{j=0}^{k} \xi_j Be(a + j, b + n - j)
$$

$\Rightarrow$ Bayes decision $\hat{\pi}^* = F_{\pi|k, \vartheta}^{-1} \left( \frac{l_1}{l_1 + l_2} \right)$
New approach

**usual burn-in strategy:** if failures occur, CM’s need to be installed. Hereafter, the burn-in study has to be repeated.

Our new approach: do not repeat burn-in, but extend the running burn-in study by increasing the sample size to $n' = n + n^*$ so that

$$\sum_{j=0}^{k} \xi_j P(X' \leq j|n', \hat{\pi}_{\text{target}} = \hat{p}_{\text{target}}) = 0.1$$

**Rationale:** Take $n^* < n$ additional devices and prove that the target ppm–level is still guaranteed on the basis of the CM model.

Efficiency of the new approach: illustration for single CM case (different degrees of effectiveness) and $k = 1, 2, 3$. 
Significant reduction of $n^*$ for high effectiveness
6. Bayesian assessment of Weibull early life failure distributions

Burn-in settings (read-outs, burn-in time, ...) are typically assessed using a Weibull DFR distribution $Wb(a, b)$ with

scale $a > 0$ and shape $b \in (0, 1)$

crucial point: joint prior $p(a, b)$

- There is no continuous conjugate joint prior
- Conjugate continuous-discrete joint prior: Gamma dist. for $a$, categorical distr. for $b$ (Soland 1969)
- Jeffreys’ prior: $p_J(a, b) \propto 1/ab$ (Sinha 1986)
We propose two alternatives:

- Histogram prior (specification remains still challenging)
- Dirichlet prior

Let $T \sim Wb(a, b)$ with density

$$f(t|a, b) \propto \begin{cases} 
 t^{b-1} \exp\left(-\left(\frac{t}{a}\right)^b\right) & t > 0 \\
 0 & \text{else}
\end{cases}$$

where $a > 0, 0 < b < 1$

Burn-in read outs at fixed time points $t_1^*, \ldots, t_m^* > 0$
Specification of the prior:

\[ F(t_i^*) \sim Be(u_i, v_i); \ u_i, v_i > 0, \ i = 1, \ldots, m \]

\[ u_i \quad \hat{=} \quad \text{prior exp. number of early life failures before time } t_i^* \]

\[ v_i \quad \hat{=} \quad \text{prior exp. number of failures surviving burn-in time } t_i^* \]

More efficiently, we summarize prior knowledge by means of a Dirichlet prior

\[ p_i = F(t_i^*) - F(t_{i-1}^*) = \text{prob. of early failure within } (t_{i-1}^*, t_i^*) \]

\[ p = (p_1, \ldots, p_{m+1})^T \sim \text{Dir}(\varphi = (\varphi_1, \ldots, \varphi_{m+1})) \]

Here we set \[ \varphi_{m+1} = \varphi^* - \sum_{i=1}^{m} \varphi_i \]
\( \vartheta^* \) regulates prior confidence through
\[
E(p) = (\vartheta_1/\vartheta^*, \ldots, \vartheta_{m+1}/\vartheta^*)
\]

Obviously:
\[
\vartheta_i = u_i - u_{i-1}; \quad i = 1, \ldots, m + 1
\]

\( \Rightarrow \) complete specification:
\[
p \sim \text{Dir}(\vartheta) \quad \text{with} \quad \vartheta = \vartheta^* E(p)
\]

**Joint prior** \( p(a, b) \) for Weibull parameters:

Draw samples of \( p_1, \ldots, p_{m+1} \) and compute

\[
F(t_i^*) = \sum_{j=1}^{i} p_j; \quad i = 1, \ldots, m
\]

Each pair \((F(t_i^*), F(t_j^*))\) with \( i, j = 1, \ldots, m; i < j \) defines a sample \((a_{ij}, b_{ij})\) of the joint prior \( p(a, b) \) via the equations
\[
F(t_i^*) = 1 - \exp\left(-\left(\frac{t_i^*}{a_{ij}}\right)^{b_{ij}}\right).
\]
\[
F(t_j^*) = 1 - \exp\left(-\left(\frac{t_j^*}{a_{ij}}\right)^{b_{ij}}\right).
\]

Explicitly, we get:

\[
b_{ij} = \frac{\ln(-\ln(1 - F(t_j^*))) - \ln(-\ln(1 - F(t_i^*))))}{\ln \frac{t_i^*}{t_j^*}}
\]

\[
a_{ij} = \exp\left\{\ln t_i^* - \frac{1}{b_{ij}} \ln(-\ln(1 - F(t_i^*)))\right\}
\]

For \(s\) Dirichlet draws \(p_1, \ldots, p_s\) we obtain \(q \times s\) pairs \((a_{ij}, b_{ij})\)

where \(q = \#\{(F(t_i^*), F(t_j^*)) : i < j = 1, \ldots, m\}\)
Whenever failures occur, the current information on the Weibull lifetime distrib. should be updated.

Data might be available as

\[ k = (k_1, \ldots, k_{m+1})^T : k_i = \# \text{failures} \in (t^*_{i-1}, t^*_i) \]

or in form of time-to-failure data

\[ t = (t_1, \ldots, t_k)^T ; \quad k = \sum_{i=1}^{m+1} k_i \]

Notice: \( k_{m+1} = \# \text{failures not detected by the burn-in is not directly available.} \)
Joint posterior \( f(a, b|k) \)

Regarding \( k = (k_1, \ldots, k_{m+1}) \) as a sample from \( MN(k, p) \), we obtain the posterior by

\[
\text{sampling } (a_{ij}|k, b_{ij}|k); \quad i = 1, \ldots, s; \quad j = 1, \ldots q
\]

according to the above equations using simulations from the Dirichlet posterior

\[ p|k \sim Dir(\vartheta + k) \]

When we are given given time-to-failure data \( t = (t_1, \ldots, t_k)^T \), then the joint posterior \( f(a, b|t) \) can be obtained according to the Metropolis-Hastings algorithm given in Kurz, Lewitschnig and Pilz (2013), where also HPD-regions for \( (a, b) \) are provided.
Update cont’d

Update of the Weibull lifetime distribution:

\[ F(t|a, b) \rightarrow F(t|\hat{a}^*, \hat{b}^*) \]

where \((\hat{a}^*, \hat{b}^*) = \arg \max_{a>0, b<1} f(a, b|\text{data})\)

\[= \text{MAP estimate} \]

Dynamical update through Bayesian learning
3 standardized read-out times

\[ t_1^* = 1 \text{h}, \ t_2^* = 2 \text{h}, \ t_3^* = 4 \text{h} \]

read-outs based on Weibull early life failure distribution

\[ T \sim Wb(a = 0.5, b = 0.75) \]

Dirichlet prior: \( E(p) = (0.81, 0.13, 0.05, 0.01) \)

expected interval failure probabilities

setting \( \nu^* = 100 \Rightarrow p \sim Dir(81, 13, 5, 1) \)
Dirichlet draws $p_1, \ldots, p_s$ define samples

$$(F(t_1^*), F(t_2^*), F(t_3^*)); \quad i = 1, \ldots, s$$

We form pairs $(F(t_1^*), F(t_2^*))$ and $(F(t_1^*), F(t_3^*))$; and proceed as shown before to get

$$\hat{a}^* = 0.505, \hat{b}^* = 0.768$$

$\Rightarrow$ prior specification is suitable

Data: $k = (k_1 = 20, k_2 = 2, k_3 = 1, k_4 = 7)^T$

$k_4 = 7$ failures not detected within $t_3^* = 4$ hours
(burn-in time not adequate)

$\Rightarrow$ posterior: $p|k \sim Dir(101, 15, 6, 8)$,

$Wb(0.5, 0.75)$ shifted to $Wb(\hat{a}^* = 0.409, \hat{b}^* = 0.485)$
HPD region for Weibull parameters \((a, b)\)

Kundu, D.: Bayesian inference and life testing plan for the Weibull distribution. *Technometrics* 50 (2008), 144-154


Kurz, D. Lewitschnig H. and Pilz J.: Decision-theoretical model for failures which are tackled by countermeasures. Submitted to *IEEE Transactions on Reliability*