

ЛЕКЦИЯ 1. Надежность вычислительных систем (ВС).

Производительность ВС. Вычислительные системы со структурно избыточностью. Стохастические модели функционирования ВС со структурно избыточностью.

Показатели надежности ВС.

Кулагин Иван Иванович

ст. преп. Кафедры вычислительных систем
Сибирский государственный университет
телекоммуникаций и информатики

Created by:

Пазников Алексей Александрович
к.т.н. доцент Кафедры вычислительных систем

- **ИВ (бакалавры):**

10 лекций, 12 лабораторных занятий, зачет

- **ИС (бакалавры):**

10 лекций, 12 лабораторных занятий, зачет

- **ИУ (бакалавры):**

10 лекций, 12 лабораторных занятий, зачет

Учебные материалы (лекции, задания на лабораторные, информация об успеваемости, литература) – на сайте:

<http://cpct.sibsutis.ru/~ikulagin/teaching/> (4 курс)

Что необходимо для оценки эффективности вычислительных систем?

1. Ввести **показатели** качества функционирования ВС.
2. Создать нетрудоёмкий и адекватный **математический аппарат** для расчёта этих показателей.
3. Разработать **технология анализа** эффективности функционирования ВС.

Под **надёжностью** (reability) понимается свойство системы сохранять заданный уровень производительности путём программной настройки её структуры и программной организации функционального взаимодействия между её ресурсами.

Современные высокопроизводительные средства обработки информации – распределённые ВС (distributed computer systems):

- Большемасштабность (large-scale), массовый параллелизм (число процессоров $\sim 10^6$)
- Программируемость структуры (structure programmability).
- Масштабируемость (scalability)
- Мультипрограммный режим.

Список TOP500

	Site	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Power (kW)
1	National University of Defense Technology China	Tianhe-2 (MilkyWay-2) - TH-IVB-FEP Cluster, Intel Xeon E5-2692 12C 2.200GHz, TH Express-2, Intel Xeon Phi 31S1P NUDT	3,120,000	33,862.7	54,902.4	17,808
2	DOE/SC/Oak Ridge National Laboratory United States	Titan - Cray XK7 , Opteron 6274 16C 2.200GHz, Cray Gemini interconnect, NVIDIA K20x Cray Inc.	560,640	17,590.0	27,112.5	8,209
3	DOE/NNSA/LLNL United States	Sequoia - BlueGene/Q, Power BQC 16C 1.60 GHz, Custom IBM	1,572,864	17,173.2	20,132.7	7,890
4	RIKEN Advanced Institute for Computational Science (AICS) Japan	K computer, SPARC64 VIIIfx 2.0GHz, Tofu interconnect Fujitsu	705,024	10,510.0	11,280.4	12,660
5	DOE/SC/Argonne National Laboratory United States	Mira - BlueGene/Q, Power BQC 16C 1.60GHz, Custom IBM	786,432	8,586.6	10,066.3	3,945

Tianhe-2 (MilkyWay-2) - TH-IVB-FEP Cluster,
Intel Xeon E5-2692 12C 2.200GHz, TH Express-2, Intel Xeon Phi 31S1P

Расположение:	Национальный университет оборонных технологий (Китай)
Производитель:	NUDT
Количество ядер:	3,120,000
Производительность Linpack (Rmax)	33,862.7 teraFLOPS
Пиковая производительность (Rpeak)	54,902.4 teraFLOPS
Электрическая мощность:	17,808.00 кВт
Память:	1,024,000 ГБ
Внутренняя сеть:	TH Express-2
Операционная система:	Kylin Linux
Компилятор:	icc
Математическая библиотека:	Intel MKL-11.0.0
MPI:	MPICH2 (GLEX channel)

Tianhe-2 (MilkyWay-2) - TH-IVB-FEP Cluster,
Intel Xeon E5-2692 12C 2.200GHz, TH Express-2, Intel Xeon Phi 31S1P



Расположение:	Национальная лаборатория Оук-Ридж (США)
Производитель:	Cray Inc.
Количество ядер:	560,640
Производительность Linpack (Rmax)	17,590.0 teraFLOPS
Пиковая производительность (Rpeak)	27,112.5 teraFLOPS
Электрическая мощность:	8,209.00 кВт
Память:	710,144 ГБ
Внутренняя сеть:	Cray Gemini interconnect
Операционная система:	Cray Linux Environment

Titan - Cray XK7 , Opteron 6274 16C 2.200GHz,
Cray Gemini interconnect, NVIDIA K20x



**Sequoia - BlueGene/Q,
Power BQC 16C 1.60 GHz, Custom**

Расположение:	Ливерморская национальная лаборатория (США)
Производитель:	IBM
Количество ядер:	1,572,864
Производительность Linpack (Rmax)	17,173.2 teraFLOPS
Пиковая производительность (Rpeak)	20,132.7 teraFLOPS
Электрическая мощность:	7,890.00 кВт
Память:	1,572,864 ГБ
Внутренняя сеть:	Custom Interconnect
Операционная система:	Linux

**Sequoia - BlueGene/Q,
Power BQC 16C 1.60 GHz, Custom**



Расположение:	Институт вычислительных систем (Япония)
Производитель:	Fujitsu
Количество ядер:	705,024
Производительность Linpack (Rmax)	10,510.0 teraFLOPS
Пиковая производительность (Rpeak)	11,280.4 teraFLOPS
Электрическая мощность:	12,659.89 кВт
Память:	1,410,048 ГБ
Внутренняя сеть:	Custom Interconnect
Операционная система:	Linux

K computer, SPARC64 VIIIfx 2.0GHz,
Tofu interconnect



Расположение:	Аргонская национальная лаборатория (США)
Производитель:	IBM
Количество ядер:	786,432
Производительность Linpack (Rmax)	8,586.6 teraFLOPS
Пиковая производительность (Rpeak)	10,066.3 teraFLOPS
Электрическая мощность:	3,945.00 кВт
Внутренняя сеть:	Custom Interconnect
Операционная система:	Linux

**Mira - BlueGene/Q,
Power BQC 16C 1.60GHz, Custom**



- Пусть N – число ЭМ в распределённой ВС.
- ВС находится в состоянии $k \in E_0^N$, $E_0^N = \{0, 1, 2, \dots, N\}$, если в ней имеется k исправных ЭМ.
- Производительность такой системы:

$$\Omega(k) = A_k k \omega$$

где ω – показатель производительности ЭМ (быстродействие по Гибсону, номинальное или среднее); A_k – коэффициент.

Что может пониматься под производительностью?

- Эффективное быстроедействие

$$\Omega^*(k) = k\omega^*$$

Число выполняемых стандартных операций (все операции, включая операции ввода/вывода).

- Среднее эффективное быстроедействие

$$\Omega(k) = \left[\sum_{i=1}^L \frac{\pi_i}{\Omega_i^*(k)} \right]^{-1}, \quad \sum_{i=1}^L \pi_i = 1$$

где $\Omega_i^*(k)$ - эффективное быстроедействие системы при решении задачи $I_i, i \in \{1, 2, \dots, L\}$, π – её вес, пропорциональный времени решения.

Вычислительные системы, komponуются, в общем случае, из не абсолютно надёжных ЭМ.

- Пусть λ - **интенсивность потока отказов** в любой из N машин.
- λ^{-1} - среднее время безотказной работы одной ЭМ (средняя наработка до отказа ЭМ)

Отказы устраняются при помощи процедуры **восстановления**

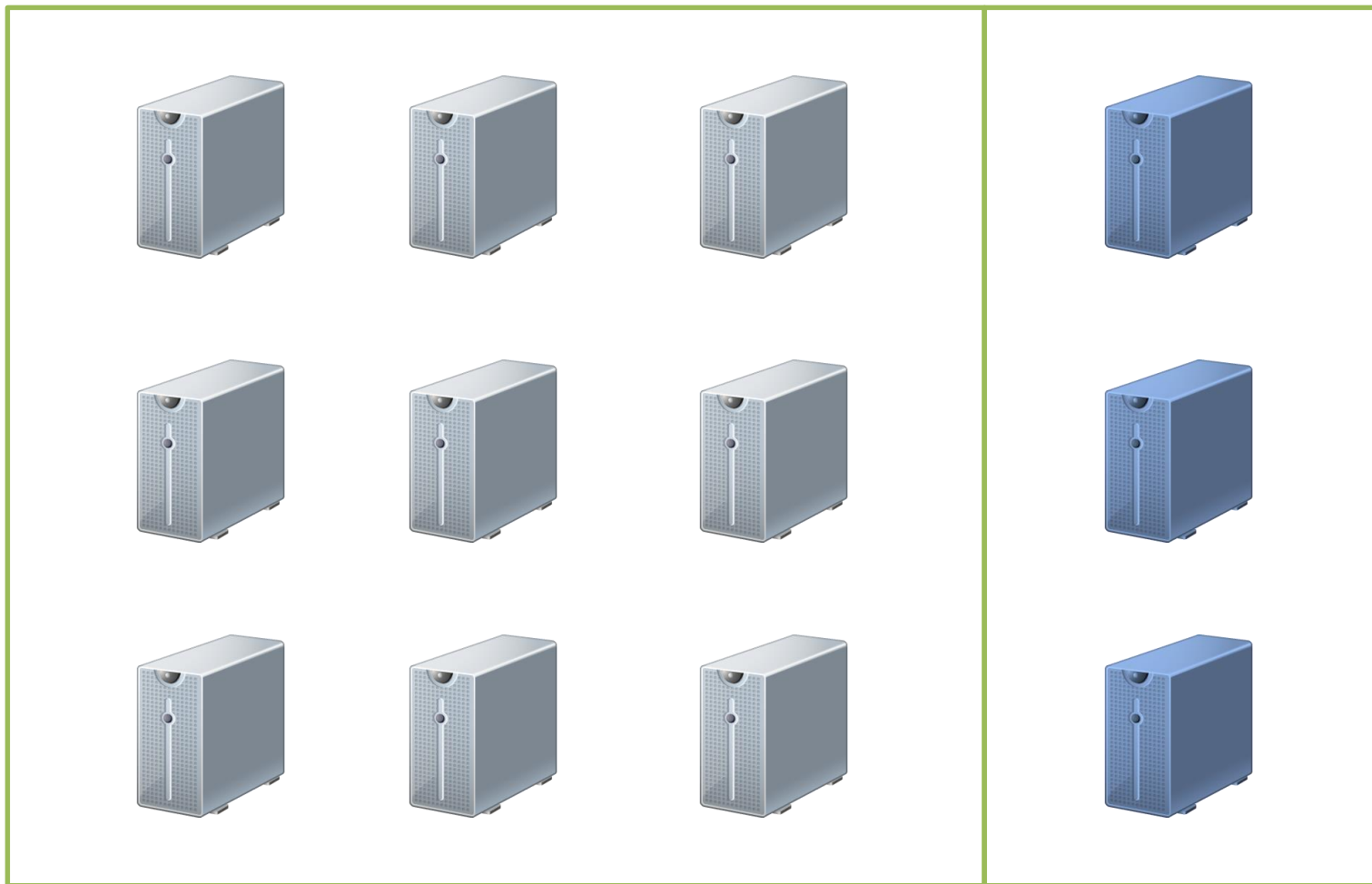
- $m, 1 \leq m \leq N$ – размер восстанавливающей подсистемы

- После отказа ЭМ поступает на восстанавливающее устройство (ВУ) или становится в очередь.
- ВУ может быть свободным и занятым восстановлением одной ЭМ.
- μ – **интенсивность восстановления**
($1/\mu$ – обнаружение + восстановление).
- В системе возможны переходы из состояния $k \in E_0^N$ в состояние $k - 1$ ($k \neq 0$) или в состояние $k + 1$ ($k \neq N$)

Вычислительные системы со структурной избыточностью

n

$N - n$



Основная
подсистема

Подчинённая
подсистема

ВС со структурной избыточностью:

1. Выделена *основная подсистема* из n ЭМ и *вспомогательная подсистема*, составляющая избыточность из $(N - n)$ машин ($n \neq 0, n \in E_0^N$).
2. Основная подсистема предназначена для решения параллельных задач из n ветвей.
3. Функции отказавшей ЭМ основной подсистемы может взять на себя исправная ЭМ вспомогательной подсистемы.

4. Производительность подчиняется следующему закону:

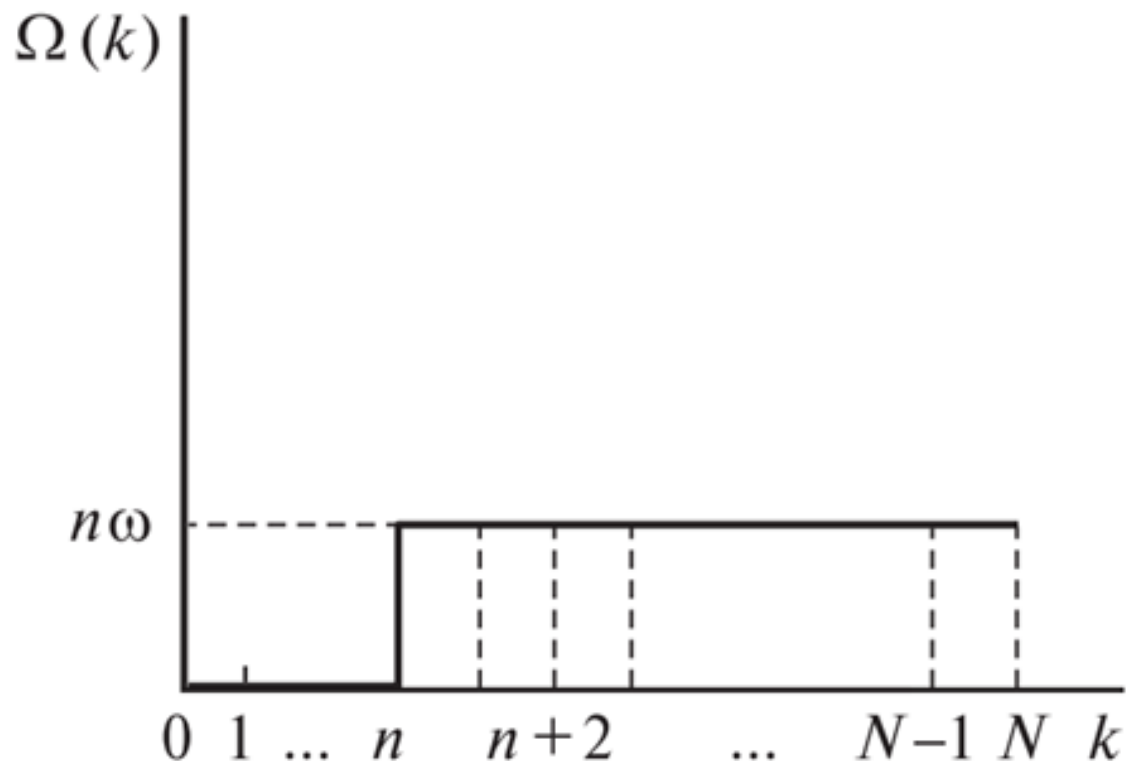
$$\Omega(k) = A_n \Delta(k - n) \varphi(n, \omega)$$

где A_n – коэффициент;

$$\Delta(k - n) = \begin{cases} 1, & \text{если } k \geq n; \\ 0, & \text{если } k < n; \end{cases}$$

ω – производительность ЭМ

$\varphi(n, \omega)$ – неубывающая функция (как правило $\varphi(n, \omega) = n\omega$)

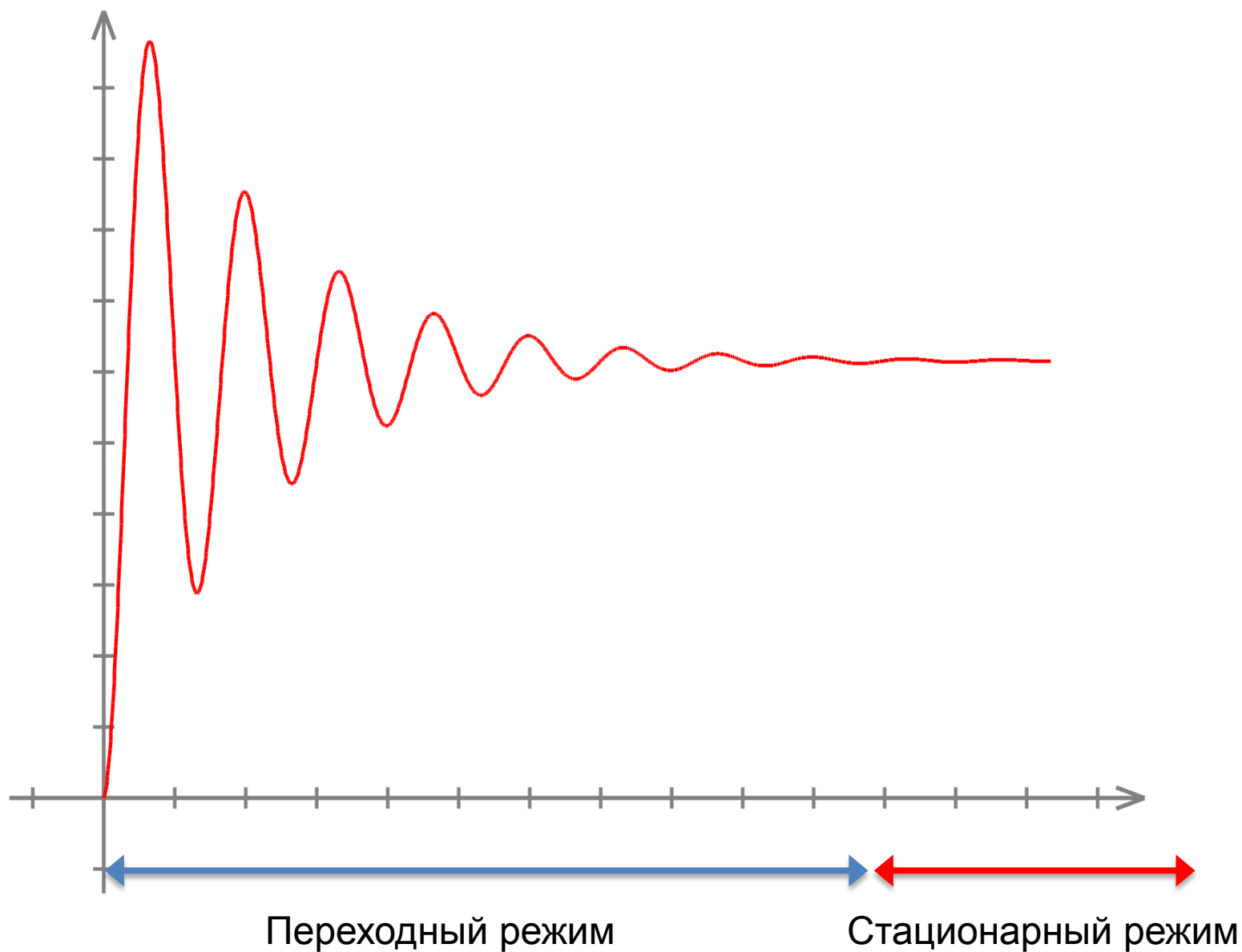


Зависимость производительности $\Omega(k)$ от k
работоспособных ЭМ для систем со структурной
избыточностью

Показатели надёжности должны характеризовать

- Производительность в текущий момент
- Производительность на промежутке времени
- Способность к восстановлению
- Функционирование ВС в переходном режиме
- Функционирование ВС в стационарном режиме

Переходный и стационарный режим



- $\xi(t)$ число исправных машин в момент времени t
- i – начальное состояние ВС ($t = 0$), $i \in E_0^N$

Функция $\xi(t)$ определяется

1. Начальным состоянием i
2. Моментами появления новых отказов
3. Моментами устранения новых отказов

Пусть известно $\xi(t^*)$ в некоторый момент времени t^*
Значения $\xi(t)$ после t^* не зависят от того, что было до t^* :

- Моменты освобождения занятых ВУ не зависят от того, что было до t^* , т.к. закон распределения времени восстановления – экспоненциальный.
- Моменты появления новых отказов не зависят от того, что было до t^* , т.к. поток отказов простейший и в нём отсутствует последствие.
- Независимость окончания устранения новых отказов также следует из экспоненциального закона.

Следовательно, $\xi(t)$ является случайным марковским процессом.

Обозначим $\{P_j(i, t)\}$ распределение вероятностей состояний системы в момент t при условии, что в начальный момент времени было исправно $i \in E_0^N$ ЭМ.

Функция $P_j(i, t)$ - вероятность того, что в системе, начавшей функционировать в состоянии $i \in E_0^N$, в момент t будет $j \in E_0^N$ исправных машин:

$$P_j(i, t) = P\{\xi(t) = j | i \in E_0^N\}, \quad j \in E_0^N$$

Тогда вероятность $P_j(i, t)$ будет показателем, характеризующим поведение ВС в **переходном режиме функционирования**.

При $i \neq j$, $i, j \in E_0^N$ имеет место:

$$P_j(i, 0) = 0, P_i(i, 0) = 1$$

Нетрудно показать, что распределение $\{P_j\}$ ($j \in E_0^N$), где

$$P_j = \lim_{t \rightarrow \infty} P_j(i, t); \quad \sum_{j=0}^N P_j = 1$$

не зависит от начального состояния $i \in E_0^N$

Следовательно, P_j ($j \in E_0^N$) - **показатель надёжности для стационарного (или установившегося) режима работы ВС.**

Для характеристики качества функционирования ВС в переходном режиме используются

- **функция надёжности $R(t)$** (вероятность безотказной работы),
- **функция восстановимости $U(t)$** (вероятность восстановления),
- **функция готовности $S(t)$.**

Функция надёжности – вероятность того, что производительность ВС, начавшей функционировать в состоянии i ($n \leq i \leq N$) на промежутке времени $[0, t)$ равна производительности основной подсистемы.

$$R(t) = P\{\forall \tau \in [0, t) \rightarrow \Omega(\tau) = A_n n \omega | n \leq i \leq N\}$$

где $\Omega(\tau)$ - производительность системы в момент времени τ .



Иначе, функция $R(t)$ есть вероятность того, что в системе, начавшей функционировать с i , $n \leq i \leq N$ исправными машинами, на промежутке времени $[0, t)$ будет не менее n исправных машин.

$$R(t) = P\{\forall \tau \in [0, t) \rightarrow \xi(\tau) \geq n \mid n \leq i \leq N\}$$

Очевидно, что $R(0) = 1$, $R(+\infty) = 0$

Функция восстановимости – вероятность того, что в ВС, имеющей начальное состояние i ($0 \leq i \leq n$), будет восстановлен на промежутке времени $[0, t)$ уровень производительности, равный производительности основной подсистемы.

$$U(t) = 1 - P\{\forall \tau \in [0, t) \rightarrow \Omega(\tau) = 0 \mid 0 \leq i < n\}$$

$$U(t) = 1 - P\{\forall \tau \in [0, t) \rightarrow \xi(\tau) < n \mid 0 \leq i < n\}$$

Очевидно, что $U(0) = 0, U(+\infty) = 1$.

В инженерной практике наиболее употребительны не $R(t)$ и $U(t)$, а математическое ожидание времени безотказной работы (средняя наработка до отказа) и среднее время восстановления:

$$\theta = \int_0^{\infty} R(t)dt$$

$$T = \int_0^{\infty} t dU(t)$$

Функция готовности – вероятность того, что производительность системы, начавшей функционировать в состоянии $i \in E_0^N$, равна в момент времени t производительности основной подсистемы:

$$S(t) = P\{\Omega(t) = A_n n \omega \mid i \in E_0^N\}$$

Иначе, функция $S(t)$ есть вероятность того, что в момент t число исправных ЭМ в ВС, имевшей начальное состояние $i \in E_0^N$, не менее числа машин основной подсистемы:

$$S(t) = P\{\xi(t) \geq n \mid i \in E_0^N\}$$

Из определения следует, что $0 < S(+\infty) < 1$,

$$S(0) = \begin{cases} 1, & \text{если } n \leq i \leq N \\ 0, & \text{если } 0 \leq i < n \end{cases}$$

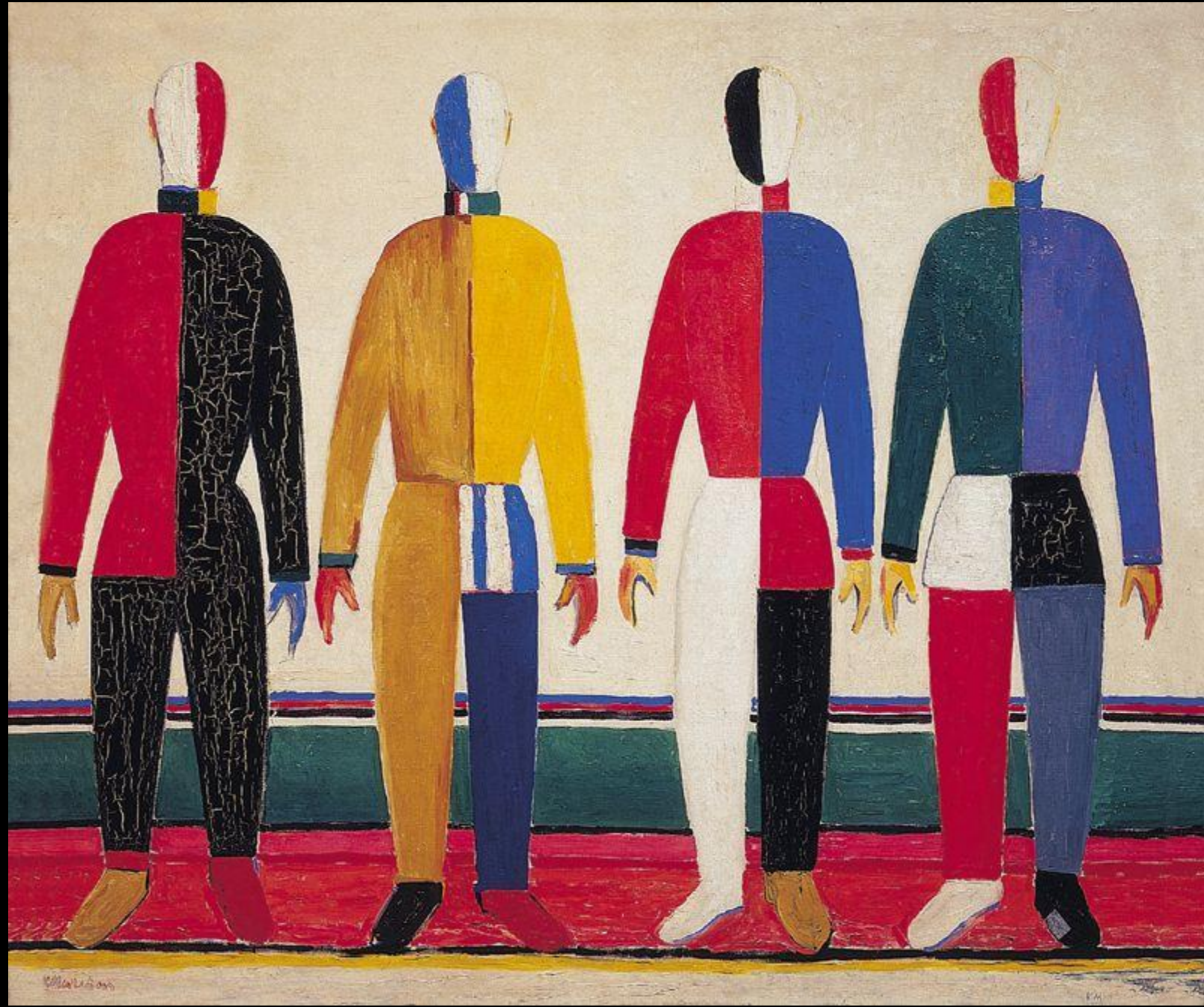
для невосстанавливаемых ВС $R(t) = S(t)$.

Функцию готовности можно выразить через вероятности её состояний:

$$S(t) = \sum_{j=n}^N P_j(i, t), \quad i \in E_0^N$$

Функция надёжности и готовности характеризуют способности ВС обеспечить требуемое быстродействие на промежутке времени $[0, t)$ и в момент t соответственно.

Функция восстанавливаемости раскрывает возможности системы к восстановлению, т.е. характеризует возможности системы к восстановлению — приобретению требуемого уровня производительности после отказа всех избыточных машин и части машин основной подсистемы.



К.Малевич. Спортсмены